Contributions to Bayesian Experimental Design

Elizabeth Ryan

BAppSc (Hons I)

under the supervision of

Professor Tony Pettitt, Dr Christopher Drovandi, and Dr M. Helen Thompson

School of Mathematical Sciences
Science and Engineering Faculty
Queensland University of Technology

2014

Submitted in partial fulfillment of the requirements of the degree of
Doctor of Philosophy

Abstract

Statistical experimental design provides rules for allocating resources in a data collection exercise where there is variability which is not under the control of the experimenter. The experimenter can use resources optimally to learn about unknowns and make predictions about future observations. Bayesian experimental design is a fast growing area of research with many real-world applications. The Bayesian framework provides a unified approach to experimental design for incorporating prior information and/or uncertainties regarding the statistical model with a utility function which describes the experimental aims.

As computational power has increased over the years, so has the development of simulation-based design methods, which involve a number of algorithms, such as Markov chain Monte Carlo. However, many of the proposed algorithms have been found to be computationally intensive for complex or nonstandard design problems, such as those which require a large number of design points to be found and/or those for which the observed data likelihood has no analytic expression.

In this thesis, we develop novel extensions of existing algorithms which have been used for Bayesian experimental design, and also incorporate methodologies which have been used for Bayesian inference into the design framework, so that solutions to more complex design problems can be found.

In Chapter 1 we discuss the aims of this research and its contributions to the literature. In Chapter 2 we provide an introduction to Bayesian experimental design and review the important literature. Chapters 3 - 5 consist of published or submitted research articles that present novel ideas for solving complex Bayesian design problems.

In Chapter 3 we discuss methods which may be used to solve design problems in which one is interested in finding a large number of (near) optimal design points (for a small number of different design variables). The approach involves the use of lower dimensional parameterisations that consist of a few design variables to generate a large number of design points which are related to one another in the design space by some (pre-specified) function. One can incorporate this approach into existing search algorithms, so that optimal values can be found for a few design variables, rather than a large number of design points. This offers substantial computational savings and a more accurate determination of the mode of the utility surface. The methodology is demonstrated on a number of different applications, including the selection of sampling times for pharmacokinetic and heat transfer studies, and involve nonlinear models.

Chapter 4 presents methods for finding fully Bayesian experimental designs for nonlinear mixed effects models. This involves the use of simulation-based optimal design methods to search over both continuous and discrete design spaces for a number of different design variables so that optimal population designs for nonlinear mixed effects models can be found. These methods have applications in the design of population pharmacokinetic studies.

Utility functions in Bayesian experimental design are based on the posterior distribution. When the posterior is found by simulation, it must be sampled from for each future

dataset drawn from the prior predictive distribution, and so many thousands of posterior distributions are often required. In Chapter 5 we compare and contrast the use of importance sampling and Laplace approximations to rapidly approximate the posterior distribution for use in Bayesian utility function calculation. These methodologies can be used to solve design problems which involve a somewhat large amount of data. Importance sampling from the prior distribution tends to break down when there is a reasonable number of experimental observations, and so Laplace approximations are used to overcome this. The methodology is motivated by a pharmacokinetic study which investigates the effect of extracorporeal membrane oxygenation on the pharmacokinetics of antibiotics in sheep. We consider several different utility functions that focus on the precise estimation of pharmacokinetic parameters/measures of interest.

In Chapter 6 we summarise this research and conclude with a discussion on the directions for future research in optimal Bayesian experimental design.

KEYWORDS: Bayesian optimal design; Decision theory; Utility function; Stochastic optimisation; Markov chain Monte Carlo; Posterior distribution approximation; Sampling strategies; Pharmacokinetics.

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Elizabeth Ryan
QUT Verified Signature

Date: 17 December, 2014

Acknowledgements

First and foremost I would like to thank all of the supervisors involved in my thesis. Firstly, I would like to thank my principal supervisor, Professor Tony Pettitt, for his support over the last few years, both academically and financially. I have benefitted a great deal from his expertise and academic guidance. Secondly, I would like to thank Dr Chris Drovandi for his invaluable supervision during my candidature. Chris has been a very supportive, patient, caring, and enthusiastic mentor. He has given me with a great deal of help with my coding, and has provided many novel ideas for my research, upon which the majority of my papers have been based. And finally, I would like to thank Dr Helen Thompson for her support during my candidature: academic, professional, and personal.

I would like to thank Dr James McGree for the help he has given me over the past few years, especially in the field of experimental design. Thank-you also to Dr Glenn Fulford for all his LaTeX help.

I would like to extend my gratitude to the academic staff in the School of Mathematical Sciences at QUT, who have been a great inspiration to me, have provided me with a solid knowledge basis and appreciation of Mathematics and Statistics, and without their early encouragement I probably would not have considered a research career in Mathematics/Statistics. I would also like to thank the School of Mathematical Sciences for their financial support which has also enabled me to attend many overseas conferences and has provided me with an invaluable learning experience. I would also like to thank QUT and the Australian government for the financial support they have given me during my candidature. This work was also supported by an APAI Scholarship which came from an ARC Linkage Grant with Roche Palo Alto (LP0991602).

A big thank-you must also go to my friends (both inside and outside of QUT) and fellow PhD students for providing me with some much needed fun and relaxation and for lending me an ear when I needed to complain about the frustrations of PhD life.

Finally, I would like to extend a MASSIVE thank-you to my family - Mum, Dad, Lou, Matt, and Anakin (my labrador). I am extremely grateful for their love, patience and support - without which I would not have been able to complete this thesis.

Contents

Chapter 1

Thesis Aims, Contributions and Outline

1.1 Description of Research Problem

Bayesian experimental design is a fast growing area of research with many real-world applications. As computational power has increased over the years, so has the development of simulation-based design methods, which involve a number of Bayesian algorithms, such as Markov chain Monte Carlo (MCMC) algorithms, that have enabled more complex design problems to be solved. In this thesis we develop novel extensions of existing Bayesian algorithms for the purpose of finding solutions to complex Bayesian experimental design problems, many of which have applications in biology and medicine. In this chapter an outline of the thesis is provided, along with its aims and the contributions of this research to the statistical literature.

1.2 Overall Objectives

The overall objective of this research is to advance Bayesian experimental design, particularly for complicated static design problems, such as those which require a large number of design variables/points to be found, and designs for mixed effects models, through the extension of existing algorithms and development of novel Bayesian methodologies and algorithms.

1.3 Specific Aims

The specific aims of this research are:

(a) To extend existing MCMC algorithms to solve static design problems which require a large number of design points to be found through the development of proposal distributions for the design points that rely on lower dimensional parameterisations of the design points;

(b) To enable fully Bayesian static designs to be found for the collection of data that is modelled by nonlinear mixed effects models;

(c) To find efficient methods for estimating the posterior distribution for use in Bayesian utility function calculation;

(d) To develop novel design criteria, based on the posterior distribution of some parameter of interest, which may be used to assist in the selection of optimal sampling times;

(e) To apply the above-mentioned methods to case study data, to improve upon the design of existing pharmacokinetic studies, demonstrate the effectiveness of the methods, and evaluate their practical benefits for the pharmaceutical and drug testing industries.

1.4 Account of Research Progress

This is a thesis by published and submitted papers. Aside from this introductory chapter and a summary in Chapter 6, the core chapters of this thesis consist of published or submitted peer-reviewed articles written primarily by the author that contribute towards solving complex Bayesian experimental design problems. These articles include three full research articles (one published and two submitted), as well as one submitted review article on Bayesian experimental design. The articles are as follows:

(a) **Ryan, E.G.**, Drovandi, C.C., McGree, J.M. and Pettitt, A.N. (2014). Fully Bayesian Experimental Design: A Review. Submitted to *International Statistical Review*. (Chapter 2).

(b) **Ryan, E.G.**, Drovandi, C.C., Thompson, M.H. and Pettitt, A.N. (2014). Towards Bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics and Data Analysis*, 70: 45-60. (Chapter 3). This paper has also been presented as a poster at the International Society for Bayesian Analysis (ISBA) World Meeting in 2012 and as a talk at the European Meeting of Statisticians (EMS) in 2013.

(c) **Ryan, E.G.**, Drovandi, C.C. and Pettitt, A.N. (2014). Simulation-based fully Bayesian experimental design for mixed effects models. Submitted to *Computational Statistics and Data Analysis*. (Chapter 4). This paper has also been presented as a poster at the International Society for Bayesian Analysis (ISBA) World Meeting in 2014.

(d) **Ryan, E.G.**, Drovandi, C.C. and Pettitt, A.N. (2014). Fully Bayesian experimental design for pharmacokinetic studies. Submitted to *Entropy*. (Chapter 5).

These journal articles are largely self-contained, but there is some degree of overlap between the literature review in Chapter 2 and the introductions of Chapters 3-5, as well as the methods used in Chapters 3-5. In particular, Chapters 4 and 5 make use of methodology introduced in Chapter 3 to reduce computational burden. The journal publications and submissions have been reprinted in full in Chapters 2-5.

1.4.1 Significance of Research

This project aims to advance Bayesian experimental design through the development of novel methodologies and use of existing Bayesian algorithms so that robust and efficient designs may be obtained for experiments that are conducted in high impact areas, such as clinical trial studies. Advancement of Bayesian statistical design will provide experimenters with appropriate statistical methodologies to conduct cost effective, timely and well informed experiments with tailored utility functions. These statistical methods will

also enhance decision making in drug development and screening by enabling experimenters to optimally allocate resources so that parameters of interest may be studied.

There is a lack of flexible and sophisticated Bayesian optimal design methodologies, particularly for complicated static design problems, such as static design problems that have a large number of design variables/points or static designs for mixed effects models. This thesis aims to address these issues by extending existing Bayesian algorithms to search for optimal designs, and by borrowing algorithms from the Bayesian inference literature to perform fast computation of the posterior distribution for Bayesian utility function calculation.

1.4.2 Thesis Structure

The structure of the thesis is as follows. In Chapter 2 we introduce the concepts involved in Bayesian experimental design and provide an extensive review of the relevant literature. In Chapter 3, we present a simulation-based approach that can be used to solve optimal design problems in which one is interested in finding a large number of (near) optimal design points using a small number of design variables. Chapter 3 addresses aims (a) and (e). Chapter 4 presents methods to find fully Bayesian designs for nonlinear mixed effects models. This chapter addresses aims (b) and (e). In Chapter 5 we compare and contrast the use of importance sampling and Laplace approximations to obtain estimates of Bayesian utility functions. Chapter 5 addresses aims (c) - (e). Finally, in Chapter 6, we summarise our methods and results, and suggest some future research that could be implemented.

Statement of Authorship for Chapter 2

This chapter has been written as a journal article. The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
- They agree to the use of the publication in the students thesis and its publication on the QUT ePrints database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for the associated publication is: **Ryan, E.G.**, Drovandi, C.C., McGree, J.M. and Pettitt, A.N. (2014). Fully Bayesian Experimental Design: A Review. Submitted to *International Statistical Review*.

Contributor	Statement of contribution
Elizabeth Ryan	Performed the literature review and wrote the manuscript, and acted as the corresponding author
Signature and Date:	
Christopher Drovandi	Directed the research and proofread the manuscript.
James McGree	Directed the research and proofread the manuscript.
Tony Pettitt	Directed the research and proofread the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all co-authors confirming their certifying authorship.

Name	Signature	Date

Chapter 2

Fully Bayesian Optimal Experimental Design: A Review

ABSTRACT

Bayesian experimental design is a fast growing area of research with many real-world applications. As computational power has increased over the years, so has the development of simulation-based design methods, which involve a number of algorithms, such as Markov chain Monte Carlo, sequential Monte Carlo and approximate Bayes methods, and which have enabled more complex design problems to be solved. This paper provides an overview of the literature on Bayesian experimental design that uses a decision-theoretic approach. The Bayesian framework provides a unified approach for incorporating prior information and/or uncertainties regarding the statistical model with a utility function which describes the experimental aims. In this paper, we provide a general overview on the concepts involved in Bayesian experimental design, and focus on describing some of the more commonly-used Bayesian utility functions and methods for their estimation, as well as a number of algorithms that are used to search over the design space to find the optimal Bayesian design. We also provide some examples from the literature of real-world applications and discuss future directions for Bayesian experimental design.

KEYWORDS: Bayesian optimal design; Decision theory; Utility function; Stochastic optimisation; Posterior distribution approximation.

2.1 Introduction

2.1.1 Background

Statistical experimental design provides rules for the allocation of resources in an information gathering exercise in which there is variability that is not under control of the experimenter. Experimental design has very broad applications across the natural, medical and social sciences, as well as engineering, business and finance. Experimental design reflects the purpose of the experiment. Prior to the commencement of an experiment, experimental design often requires choices to be made regarding which treatments to study and how these treatments will be defined or administered (e.g., amount, timing, frequency), and the proportion of observations to allocate to each treatment. Experimental design can also require choices of blocking factors, randomisation methods and sample size to be made. The experimental units must also be clearly defined prior to the commencement of the experiment, along with the time period over which the experiment

is to be performed. Due to costs, ethics and other constraints on time, efficient use of resources is highly critical.

Experimental designs incorporate features into studies with the aim to control systematic error (bias), reduce random variations, increase precision of parameter estimates (or some measure of interest), make predictions about future observations, or discriminate between competing models. Essentially, non-optimal designs require more resources to make inferences on the features of interest with the same level of reward that an optimal design would. Experimental design problems are commonly viewed as optimisation problems, and optimal experimental designs may be used to achieve the experimental goals more rapidly and hence reduce experimental costs.

Experimental design has been widely developed within the classical framework, in both theory and practice (e.g., Atkinson and Donev (1992)). In the classical framework, optimal experimental designs are commonly derived using optimality criteria that are based on the expected Fisher information matrix (e.g., Fedorov (1972); Pukelsheim and Torsney (1991); Atkinson and Donev (1992)).

Classical experimental design is well suited to linear or linearised models. For nonlinear models, optimal designs generally depend on the true values of the model parameters (assuming the model is also true). Often, the aim of experimental design is to precisely estimate model parameters. Since the parameter values are not known, and data has not been collected to estimate them, the experimenter must postulate values for the model parameters from which to construct an experimental design. Use of unlikely parameter values may result in sub-optimal designs. Several studies have incorporated probability distributions on the model parameters and averaged local design criteria over the distributions so that the designs obtained may be robust to the initial choice of the parameter values (e.g., Pronzato and Walter (1985); D'Argenio (1990)). These probability distributions are known as *prior distributions* and can incorporate information from previous studies, expert elicited data or subjective beliefs of the experimenters. Similar methods are also used for situations in which there is model uncertainty.

Bayesian statistics is rapidly gaining in popularity in the literature and has many applications, particularly in the fields of science, health and engineering. Bayesian statistics combines prior knowledge about the unknown parameters in the model with the likelihood (contribution made by the data to the unknown parameters) to give the posterior distribution, from which inferences on the unknown parameters of interest can be made.

Designs which have arisen from averaging classical design criteria over prior distributions have commonly been referred to as "Bayesian designs". We suggest this is misleading as we propose that to qualify as a "fully Bayesian design", one must obtain the design by using a design criterion that is a functional of the posterior distribution. Designs which have arisen from averaging the classical design criteria over the parameter space are termed "pseudo-Bayesian" or "robust" designs (Pronzato and Walter (1985); Federov and Hackl (1997)).

Bayesian methodologies for optimal experimental design have become more prominent in the literature (e.g., Müller (1999); Han and Chaloner (2004); Amzal et al. (2006); Müller et al. (2006); Cook et al. (2008); Huan and Marzouk (2013)). One advantage of using a Bayesian design criterion is that a single design point can be used, and the prior distribution is updated by the single observation. Lindley (1972) presents a decision theoretic approach to experimental design, upon which Bayesian experimental design is based. Bayesian optimal design involves defining a design criterion, or a utility function $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$, that describes the worth (based on the experimental aims) of choosing the design $\mathbf{d}$ from the design space $\mathbf{D}$ yielding data $\mathbf{y}$, with model parameter values $\boldsymbol{\theta}$. A probabilistic model, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$, is also required. This consists of a likelihood $p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta})$ for observing a new set of measurements $\mathbf{y}$ at the design points $\mathbf{d}$, given parameter values $\boldsymbol{\theta}$, and a prior distribution $p(\boldsymbol{\theta})$ for the parameters $\boldsymbol{\theta}$. The prior distribution is usually assumed to be independent of the design $\mathbf{d}$.

The Bayesian optimal design, $\mathbf{d}^*$, maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$ with respect to the future data $\mathbf{y}$ and model parameters $\boldsymbol{\theta}$:

$$\begin{aligned}
\mathbf{d}^* &= \arg\max_{\mathbf{d}\in\mathbf{D}} E\{U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})\} \\
&= \arg\max_{\mathbf{d}\in\mathbf{D}} \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \\
&= \arg\max_{\mathbf{d}\in\mathbf{D}} \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{y}. \qquad (2.1)
\end{aligned}$$

Thus, the optimal design (given the observed data), maximises the posterior expected utility. Unless the likelihood and prior are specifically chosen to enable analytic evaluation of the integration problem, equation (2.1) does not usually have a closed form solution. Therefore, numerical approximations or stochastic solution methods are required to solve the maximisation and integration problem.

Due to the computational challenges of performing the integration and maximisation of equation (2.1), the use of standard optimisation algorithms, such as the Newton-Raphson method, to find the optimal design is inappropriate. A number of stochastic algorithms have been proposed in the literature to approximate the maximisation and integration problem of equation (2.1). These include: prior simulation (Müller (1999)); smoothing of Monte Carlo simulations (Müller (1999)); gridding methods which involve numerical quadrature or Laplace approximations to perform backward induction (Brockwell and Kadane (2003)); Markov chain Monte Carlo simulation in an augmented probability model (Müller (1999)); and sequential Monte Carlo methods (Kück et al. (2006); Amzal et al. (2006)). These algorithms will be discussed further in Sections 2.5 and 2.6.

2.1.2 Brief History of the Bayesian Design Literature

A broad range of literature exists on optimal experimental design. This article aims to review those papers which present solutions to fully Bayesian experimental design problems.

Early work on Bayesian decision-theory includes Lindley (1968, 1972), which notes that the design of an experiment should depend on the experimental objectives (e.g., precise estimation of certain parameters, prediction of future responses). Other work includes Chaloner (1984), who further developed Bayesian optimal design theory in a linear regression context and explicitly describes how the prior causes a difference between the classical and Bayesian optimal designs. DasGupta and Studden (1991) gave a structured formulation to demonstrate the sensitivity of designs to the priors and presented designs that were robust to the prior specification. Other notable works on Bayesian designs for linear regression models include Pilz (1991) and El-Krunz and Studden (1991). Simulation-based design methods have frequently been used more recently (e.g., Clyde et al. (1996); Bielza et al. (1999); Müller (1999); Stroud et al. (2001); Amzal et al. (2006); Müller et al. (2006); Cook et al. (2008); Cavagnaro et al. (2010)) in which Markov chain Monte Carlo and sequential Monte Carlo algorithms are utilised to solve complex optimal Bayesian design problems (e.g., designing for nonlinear models). Sequential, or adaptive designs, have become increasingly popular in the Bayesian design literature as they provide flexible and efficient designs. Rather than using the same design throughout the experimental process, as in *static* design problems, the design which maximises the expected utility is chosen at each stage of experimentation, based on the outcomes of previous experiments. Recent developments in static and sequential designs will be discussed further in Sections 2.5 and 2.6.

There are already several notable review papers on Bayesian experimental design. DasGupta (1995) presents a review of both classical and Bayesian experimental design, with a focus on designing for linear models. Atkinson (1996) review classical and pseudo-Bayesian optimal design for linear and nonlinear models. Verdinelli (1992) and Chaloner and Verdinelli (1995) present a comprehensive review on Bayesian experimental design, for both linear and nonlinear models. Müller (1999) provides an overview of simulation-based methods in optimal design. Clyde (2001) presents a broad review on several of the key concepts involved in Bayesian experimental design, such as, choice of utility functions; prior elicitation; and methods for calculating the expected utility.

2.1.3 Contribution and Outline

There has been a lack in review papers on fully Bayesian experimental design since the early 2000s. These earlier review papers have often been written from a rather mathematical view point, and have often focused on defining Bayesian design criteria and their relationship to classical design criteria. In the past two decades there has been a substantial increase in computational power and, along with it, the use of Bayesian methodologies for optimal design. At the present time, we have been unable to find any recent review articles which discuss the various algorithms that are used in the Bayesian design literature to solve optimal design problems. Designs for complex models have also received little attention in Bayesian experimental design literature reviews. This article is concerned with reviewing the computational methods that have been used to find fully Bayesian experimental designs and aims to address the aspects of Bayesian experimental

design which have received little or no emphasis in previous review papers. This article is aimed at readers with some understanding of Bayesian methods, but not necessarily with knowledge of experimental design.

In Section 2.2 we describe how the prior distribution has been elicited in previous Bayesian experimental design studies. Section 2.3 discusses methods for posterior distribution approximation for use in Bayesian utility functions. In Section 2.4 we discuss some of the more commonly-used Bayesian utility functions, along with the methods that have been used for their estimation. Sections 2.5 and 2.6 provide an overview of the optimisation algorithms that have been used to search for static and sequential Bayesian experimental designs, respectively.

2.2 The Prior Distribution

The Bayesian design framework (as well as Bayesian analysis) requires the elicitation of a prior distribution for the statistical model(s). It is very important to check the sensitivity of the optimal design to the specification of the prior distribution (see, for example, DasGupta and Studden (1991)).

A number of studies (e.g., Clyde et al. (1996); Stroud et al. (2001); Ryan et al. (2014a)) have used historical data from previous experiments to construct a prior distribution for the design of future experiments. Toman and Gastwirth (1994) suggest the use of results from a pilot study to specify the prior distribution. Tsai and Chaloner (2002) used information from over 50 clinical experts to elicit prior distributions for their design problem. Kadane (1996) discusses a number of the practical issues that occur in subjective elicitation for clinical trials.

Several studies have considered the problem where the prior used for the design phase is different from the prior used for analysis (e.g., Etzione and Kadane (1993); Han and Chaloner (2004)). For example, a noninformative prior may be used for analysis to mimic a classical analysis, but all available prior information may be used in the design process so that an informative prior may be used.

2.3 Estimation of the Posterior Distribution

Bayesian utility functions are based on the posterior distribution and generally assume that a Bayesian analysis will be performed on any data that are generated from the experimental design. In general, the posterior distribution does not have a closed form expression, and numerical methods are required to sample from or approximate the posterior distribution. Generally, thousands of posterior distributions need to be considered, since each possible future data set that is drawn from the prior predictive distribution requires calculations of the posterior distribution. For this to be computationally feasible, rapid methods for obtaining the posterior distributions for many datasets that are drawn from the same model and prior are required.

2.3.1 Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) has often been used to estimate the posterior distribution for Bayesian utility function calculations (e.g., Wakefield (1994); Palmer and Müller (1998); Han and Chaloner (2004)). Although MCMC is often appropriate and useful for Bayesian data analysis, it can be too computationally intensive to perform MCMC to estimate the posterior distribution for each of the thousands of iterations required in the Bayesian experimental design algorithms.

2.3.2 Importance Sampling

Importance sampling is a popular method for estimating target distributions of interest, from which it may be difficult to sample (Geweke (1989)). Importance sampling involves choosing an *importance distribution* $g(\cdot)$, from which it is easy to sample, and then appropriately weighting the samples that have been drawn from the importance distribution to account for the discrepancy between $g(\cdot)$ and the target distribution. In the Bayesian design context, the target distribution is the posterior $p(\boldsymbol{\theta}|\mathbf{d},\mathbf{y})$. Weighted samples $\{\boldsymbol{\theta}_k, W_k\}_{k=1}^{N_p}$ are produced, where N_p is the number of particles used to estimate the posterior; $w(\boldsymbol{\theta}) = \frac{p(\mathbf{y}|\mathbf{d},\boldsymbol{\theta})p(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}$ are the importance weights; and $W_k \propto w(\boldsymbol{\theta}_k), k = 1, ..., N_p$ are the normalised importance weights, $\sum_{k=1}^{N_p} W_k = 1$. The target and importance distributions should have the same support. To measure the efficiency of importance sampling, the effective sample size (ESS) is used and can be approximated via

$$ESS = \frac{1}{\sum_{k=1}^{N_p} W_k^2}, 1 \leq ESS \leq N_p.$$

Importance sampling is a very useful method for estimating the posterior distribution in Bayesian experimental design since the importance samples only need to be drawn once (unlike MCMC) and can then be re-weighted in each iteration of the optimisation algorithm according to the current design and data. The ability to re-use the importance samples offers substantial computational savings.

Importance sampling from the prior distribution has commonly been used in Bayesian experimental design to estimate the posterior distribution (e.g., Cook et al. (2008); McGree et al. (2012c); Ryan et al. (2014a,c)). This reduces the importance weights to be proportional to the likelihood function. However, this is usually inefficient when there is a substantial difference between the prior and posterior distributions (e.g., Bengtsson et al. (2008); Ryan et al. (2014a,c)).

Ryan et al. (2014a) used Laplace approximations (to the posterior) to form the importance distribution for importance sampling, and found that this approach corrects for some non-normality that is not accommodated by the Laplace approximation, and can also be used when large amounts of data are involved in the design problem since fewer particles are required in the importance sampling to obtain a reasonable ESS.

The use of adaptive importance sampling (e.g., Kinas (1996); Pennanen and Koivu (2006)) is largely unexplored for estimating the posterior distribution in Bayesian experimental design problems.

2.3.3 Deterministic Approximations

Laplace approximations (or Gaussian approximations) and numerical quadrature provide fast methods for obtaining approximations to the posterior distribution in Bayesian design problems (e.g., Lewi et al. (2009); Cavagnaro et al. (2010); Bornkamp et al. (2011); Long et al. (2013); Ryan et al. (2014a)). These methods are particularly useful when large amounts of data are involved. However, their suitability depends on whether it is reasonable to assume that the posterior distribution is well approximated by a multivariate normal distribution and they also suffer from the curse of dimensionality. To overcome the issue of dimensionality, Long et al. (2013) use polynomial-based sparse quadrature for the integration over the prior distribution.

Integrated nested Laplace approximation (INLA) is a relatively new method for rapidly approximating posterior distributions (see Rue et al. (2009)). INLA generally is a significantly faster alternative to MCMC and importance sampling for approximating the posterior. To date, INLA has mostly been used for approximate posterior inference for models in which the posterior marginals are not available in closed form due to non-Gaussian response variables, such as latent Gaussian Markov random field (GMRF) models (e.g., Rue et al. (2009)) with non-Gaussian observations. INLA enables fast Bayesian inference by using accurate approximations to the marginal posterior density for the hyperparameters and the posterior marginal densities for the latent variables. The use of INLA in the context of Bayesian experimental design is currently unexplored.

Variational Bayesian (VB) methods facilitate approximate inference for intractable posteriors (or other target densities) and provide an alternative to other approaches for approximate Bayesian inference such as MCMC and Laplace approximations. VB can also be used to determine a lower bound for the evidence for use in model selection problems. The VB approach is fast and deterministic, and involves approximating the intractable target densities, e.g., $p(\boldsymbol{\theta}|\mathbf{y})$, by a factored form $q(\boldsymbol{\theta}) = q_1(\boldsymbol{\theta}_1) \times .... \times q_r(\boldsymbol{\theta}_r)$, for which $q(\boldsymbol{\theta})$ is more tractable than $p(\boldsymbol{\theta}|\mathbf{y})$. An issue is the factorization for the variational approximation $q(\cdot)$. Variational approximations have commonly been used for Bayesian inference (e.g., Ormerod and Wand (2010)), but have not yet been used in a Bayesian experimental design context. These methods could provide a fast alternative for approximating the posterior for use in Bayesian utility function calculation. However, the error of the VB approximation is generally unknown and can be substantial (e.g., Rue et al. (2009)).

2.3.4 Approximate Bayesian Computation

Approximate Bayesian computation (ABC) is a likelihood-free method that is used to approximate the posterior distribution in situations where the likelihood function is intractable, but simulation from the likelihood is relatively straightforward. ABC has commonly been used to perform inference (e.g., Drovandi and Pettitt (2011); Drovandi et al.

(2011); Sisson and Fan (2011)). One of the most common ABC algorithms is ABC rejection (see Beaumont et al. (2002)). ABC rejection prevents one from having to evaluate the likelihood by instead drawing many parameter values from the prior, and simulating data from the model, conditional on those parameter values. Only those parameters that generate simulated data that are close in some sense to the observed data are kept. The efficiency of this method is dependent on how close the posterior distribution is to the prior.

Drovandi and Pettitt (2013) and Hainy et al. (2013) used ABC rejection in the Bayesian experimental design context to approximate the posterior distributions (for Bayesian utility function calculation) for models with computationally intractable likelihoods. The ABC posterior is given by:

$$p(\boldsymbol{\theta}|\mathbf{d},\mathbf{y},\epsilon) = \int_{\mathbf{x}} p(\mathbf{x}|\mathbf{d},\boldsymbol{\theta})p(\boldsymbol{\theta})1(\rho(\mathbf{y},\mathbf{x}) \leq \epsilon)d\mathbf{x},$$

where $\mathbf{y}$ represents the 'observed data' (that is generated from the model at each iteration of the optimisation (e.g., MCMC) algorithm); $\mathbf{x}$ is simulated data; $1(\cdot)$ is an indicator function; $\rho(\cdot,\cdot)$ is function that measures the discrepancy between the observed and simulated data; and ϵ is a tolerance threshold that controls the error of the approximation. The discrepancy function typically compares summary statistics of the observed and simulated data. However, Drovandi and Pettitt (2013) only considered low dimensional designs and so they were able to compare the observed and simulated data directly. ABC rejection is very useful since the ABC data, i.e., the $\mathbf{x}$ values, as well as the model parameters $\boldsymbol{\theta}$, only need to be simulated once and can be re-used at each iteration of the optimisation algorithm (much in the same spirit as importance sampling) for comparison to the observed data, $\mathbf{y}$. This offers substantial computational savings.

2.4 Bayesian Utility Functions and Methods for Their Estimation

It is highly important that the utility function incorporates the experimental aims and is specific to the application of interest. For instance, designs which efficiently estimate the model parameters may not be useful for prediction of future outcomes. Several approaches have been suggested in the literature to assist in the elicitation of the utility function (see Spiegelhalter et al. (1996); Wolfson et al. (1996)). In practice, the utility function is often not specified as a single function, due to the difficulty of combining competing goals, and instead a set of possible utility functions is used. Christen et al. (2004) formally acknowledged the fact that the decision maker may be unwilling or unable to specify a unique utility function by considering a set of possible utility functions. Sensitivity analyses to misspecifications in the utility function have been proposed (see Rios Insua and Ruggeri (2000) for a review). In this section we will discuss some of the more commonly used Bayesian utility functions, as well as methods for their estimation based on the approximation to the posterior. One of the most commonly used and versatile Bayesian design criteria is the mutual information, which is based on entropy, and has been used for designing for efficient parameter estimation (Bernardo (1979); Ryan (2003);

Paninski (2005)), as well as minimising prediction uncertainty (Liepe et al. (2013)), and model discrimination (Box and Hill (1967); Ng and Chick (2004); Cavagnaro et al. (2010); Drovandi et al. (2014)). For discussion of other Bayesian utility functions, see Chaloner and Verdinelli (1995).

For normal linear models, analytical expressions for equation (2.1) can be obtained for many Bayesian utilities, provided the model dimension and decision space is small (e.g., Borth (1975); Chaloner and Verdinelli (1995); Ng and Chick (2004)). For nonlinear design problems, one cannot usually obtain an analytical expression, and the integrals in equation (2.1) can instead be approximated by Monte Carlo methods (e.g., Palmer and Müller (1998); Cook et al. (2008); Ryan et al. (2014a)), Laplace approximations (e.g., Lewi et al. (2009); Ryan et al. (2014a)), or numerical quadrature (e.g., Cavagnaro et al. (2010)).

2.4.1 Parameter Estimation Utility Functions

Precise parameter estimation is a common goal of experimental design and many different utility functions have been used to achieve this purpose. Bayesian utility functions that design for precise parameter estimation are discussed below.

Information-based Utilities

When interest lies in estimating some function of $\boldsymbol{\theta}$, say $\phi(\boldsymbol{\theta})$, the mutual information between $\phi(\boldsymbol{\theta})$ and the data $\mathbf{y}$, conditional on the design $\mathbf{d}$, may be given by:

$$\begin{aligned} I(\phi(\boldsymbol{\theta}); \mathbf{y}|\mathbf{d}) &= U(\mathbf{d}) \\ &= \int_{\phi(\boldsymbol{\theta})} \int_{\mathbf{Y}} p(\phi(\boldsymbol{\theta}), \mathbf{y}|\mathbf{d}) \Big[\log p(\phi(\boldsymbol{\theta}), \mathbf{y}|\mathbf{d}) - \log p(\mathbf{y}|\mathbf{d}) - \log p(\phi(\boldsymbol{\theta})) \Big] d\mathbf{y} d\phi(\boldsymbol{\theta}). \end{aligned} \tag{2.2}$$

The optimal design that maximises the utility function is the one that yields the largest information gain, on average, about $\phi(\boldsymbol{\theta})$ upon observation of the data.

Another commonly-used Bayesian design criterion is the Kullback-Leibler divergence (KLD) (Kullback and Leibler (1951)) between the prior and posterior distributions, which is given by:

$$\begin{aligned} U(\mathbf{d}, \mathbf{y}) &= E_{\phi(\boldsymbol{\theta})|\mathbf{d},\mathbf{y}}(\log p(\phi(\boldsymbol{\theta})|\mathbf{d}, \mathbf{y}) - \log p(\phi(\boldsymbol{\theta}))) \\ &= \int_{\phi(\boldsymbol{\theta})} p(\phi(\boldsymbol{\theta})|\mathbf{d}, \mathbf{y}) \log p(\mathbf{y}|\mathbf{d}, \phi(\boldsymbol{\theta})) d\phi(\boldsymbol{\theta}) - \log p(\mathbf{y}|\mathbf{d}). \end{aligned} \tag{2.3}$$

Lindley (1956) suggested that this utility should be used if one is interested in maximising the expected information gain on the model parameters (or functions of) due to performing an experiment at design points $\mathbf{d}$. Mathematically, the mutual information is the KLD between the joint distribution $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ and product of marginal distributions of $\boldsymbol{\theta}$ and $\mathbf{y}$ (Borth (1975)). This criterion is equivalent to the classical D-optimal criterion when designing for normal linear models with a normal prior distribution for the model parameter (see Chaloner and Verdinelli (1995) and Verdinelli (2000) for further details).

Ryan (2003) used mutual information to find static designs for efficient parameter estimation. Kim et al. (2013) used the mutual information utility to find sequential designs to efficiently estimate parameters, which was of the form:

$$U(\mathbf{d}_{(t)}) = \int_{\boldsymbol{\Theta}} \int_{\mathbf{Y}} \left[\log \left(\frac{p(\boldsymbol{\theta}|\mathbf{d}_{(t)}, \mathbf{y}_{(1:t)})}{p(\boldsymbol{\theta}|\mathbf{y}_{(1:t-1)})} \right) \right] p(\mathbf{y}_{(t)}|\mathbf{d}_{(t)}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}_{(1:t-1)}) d\mathbf{y}_{(t)} d\boldsymbol{\theta},$$

where $\mathbf{y}_{(1:t)}$ are the data that were observed from the 1st to the t-th trial, $\mathbf{y}_{(t)}$ are the data that were observed at the current, t-th trial, using design $\mathbf{d}_{(t)}$, $\mathbf{y}_{(1:t-1)}$ are the data that were measured from the 1st to the $(t-1)$-th trials using the designs $\mathbf{d}_{(1:t-1)}$. Paninski (2005) proved that under acceptably weak modelling conditions, utility functions based on mutual information can choose designs that lead to consistent and efficient parameter estimates in the adaptive design framework.

Despite the theoretical appeal, mutual information is computationally complex, due to the difficulty in calculating the evidence $p(\mathbf{y}|\mathbf{d})$ in equation (2.2). Therefore, many design problems have been restricted to special cases, such as designing for parameter estimation of linear gaussian models (e.g., Lewi et al. (2009)) or binary models (e.g., Kujala and Lukka (2006)) in which the evidence can be computed analytically. Conjugate priors have been used to obtain analytic results (e.g., Borth (1975)) and numerical quadrature has also been used (e.g., Cavagnaro et al. (2010)). Drovandi et al. (2013) used sequential Monte Carlo algorithms (which are described in more detail in Section 2.5) for both posterior and evidence approximation so that the mutual information could be calculated for sequential design problems for parameter estimation. Ryan et al. (2014c) used importance sampling to calculate the KLD between the prior and posterior distributions for static design problems, but found this to be computationally intensive. Huan and Marzouk (2012, 2013) used polynomial chaos approximations and nested Monte Carlo integration (Ryan (2003)) to estimate the KLD between the prior and posterior distributions for static design problems for parameter estimation.

Scalar Functions of the Posterior Covariance Matrix

The inverse of the determinant of the posterior covariance matrix is a useful utility function if one is interested in maximising the (joint) posterior precision of all (or a subset) of the model parameters $\boldsymbol{\theta}$ (e.g., Drovandi et al. (2013); Ryan et al. (2014c)) or a function of the model parameters $\phi(\boldsymbol{\theta})$ (e.g., Stroud et al. (2001); Drovandi et al. (2013); Ryan et al. (2014a)). This utility is also known as the "Bayesian D-posterior precision" and is given by:

$$U(\mathbf{d}, \mathbf{y}) = \frac{1}{\det(\text{cov}(\phi(\boldsymbol{\theta})|\mathbf{d}, \mathbf{y}))}.$$

If one were interested in maximising the precision of the marginal posterior distributions of the model parameters, then one should use the trace instead of the determinant to obtain the Bayesian A-posterior precision. The Bayesian D-posterior precision is much less computationally intensive to estimate than equation (2.2). However, if the posterior

distribution is multi-modal, then use of the Bayesian D-posterior precision utility may be inappropriate and one should instead use equation (2.2) as the utility function.

The posterior variance-covariance matrix can easily be obtained from the weighted posterior samples that are obtained from importance sampling (e.g., Stroud et al. (2001); McGree et al. (2012c); Ryan et al. (2014a)), ABC rejection (e.g., Drovandi and Pettitt (2013)) and via sequential Monte Carlo (Drovandi et al. (2013)). The posterior variance-covariance matrix is also easily obtained when one uses numerical quadrature or Laplace approximations to the posterior distribution.

Quadratic Loss

When one is interested in obtaining a point estimate of the parameters, or linear combinations of them, a quadratic loss function may provide a suitable utility function:

$$U(\mathbf{d},\mathbf{y}) = -\int_{\phi(\boldsymbol{\theta})} (\phi(\boldsymbol{\theta}) - \widehat{\phi(\boldsymbol{\theta})})^T \mathbf{A} (\phi(\boldsymbol{\theta}) - \widehat{\phi(\boldsymbol{\theta})}) p(\phi(\boldsymbol{\theta})|\mathbf{d},\mathbf{y}) d\phi(\boldsymbol{\theta}),$$

where $\mathbf{A}$ is a symmetric non-negative definite matrix (e.g., Chaloner (1984); Chaloner and Verdinelli (1995); Han and Chaloner (2004)) and $\widehat{\phi(\boldsymbol{\theta})}$ is some estimate (e.g., the mean) of $p(\phi(\boldsymbol{\theta})|\mathbf{d},\mathbf{y})$. Once the posterior distribution has been approximated, it is quite straightforward to estimate this utility. For normal linear models, when one is interested in point estimates of parameters, this utility is the Bayesian equivalent of the classical A-optimal criterion. When one is interested in linear combinations of the parameters, this utility is the Bayesian equivalent of the classical C-optimal criterion (for normal linear models).

2.4.2 Utilities for Model Discrimination

Model discrimination is an important experimental design problem which has generated a substantial amount of research (see, for example, Box and Hill (1967); Hill et al. (1968); Borth (1975); Cavagnaro et al. (2010); Drovandi et al. (2014)). Much of the design literature has focused on producing designs that offer efficient and precise parameter estimates. However, these designs can perform poorly on model discrimination problems (see, for example Atkinson (2008); Waterhouse et al. (2009)).

Mutual information has commonly been used as the utility function in the Bayesian design literature to design for model discrimination (e.g., Box and Hill (1967); Borth (1975); Ng and Chick (2004); Cavagnaro et al. (2010); Drovandi et al. (2014); McGree et al. (2012b)). The optimal design $\mathbf{d}$ is the one that maximises the mutual information between the (random variable) model indicator, m, and the future observation $\mathbf{y}$ (see, for example, Cavagnaro et al. (2010)). Drovandi et al. (2014) give an expression of this utility to design for model discrimination for discrete data, and McGree et al. (2012b) provide an expression for continuous data. Both Drovandi et al. (2014) and McGree et al. (2012b) used sequential Monte Carlo methods to approximate the necessary quantities so that mutual information could be used to obtain sequential designs for model discrimination.

Roth (1965) proposed a model discrimination utility that is known as 'total separation', and selects design points that yield the largest differences between the posterior predictive means of rival models. This is achieved by maximising a weighted sum (over all of the potential models) of the product of the absolute differences between the posterior predicted mean responses from all rival models and the given ('true') model. Total separation has recently been used by Masoumi et al. (2013) and McGree et al. (2012b) to design for model discrimination. The total separation utility can be approximated quite easily once the posterior predictive distribution has been found (see, for example McGree et al. (2012b)). This utility does not account for the variance of the predicted responses (Hill (1978)), which is problematic if the competing models differ in their error structures (e.g., additive vs. multiplicative error) (McGree et al. (2012b)).

Both mutual information and total separation do not rely on the assumption of a particular model being true (unlike many of the classical design criteria), but require the experimenter to define a set of rival models with prior probability of being true. That is, these utilities use the M-closed approach of Bernardo and Smith (2000, chapter 6).

Vanlier et al. (2014) proposed a model discrimination utility that is based on a k-nearest neighbour estimate of the Jensen Shannon divergence (which is the averaged KLD between the probability densities and their mixture) between the multivariate predictive densities of competing models. They showed that their utility is monotonically related to the expected change in the Bayes Factor in favour of the model that generated the data. MCMC was used to sample from the posterior distributions and the predictive distributions were sampled using these posterior distribution values and by adding noise generated by the error model. This was found to be computationally intensive, especially for their application which involved nonlinear models of biochemical reaction networks.

2.4.3 Utilities for Prediction of Future Observations

If one is interested in choosing $\mathbf{d}$ to predict $\mathbf{y}_{n+1}$ from $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$, then the expected gain in Shannon information for a future observation, $\mathbf{y}_{n+1}$, from the prior predictive distribution to the posterior predictive distribution can be used as the utility function:

$$U(\mathbf{d}_{n+1}, \mathbf{y}) = \int_{\boldsymbol{\theta}} \int_{\mathbf{Y}_{n+1}} p(\mathbf{y}_{n+1}|\mathbf{d}_{n+1}, \mathbf{y}_{1:n}, \boldsymbol{\theta}) \log p(\mathbf{y}_{n+1}|\mathbf{d}_{n+1}, \mathbf{y}_{1:n}, \boldsymbol{\theta}) d\mathbf{y}_{n+1} d\boldsymbol{\theta} - \log p(\mathbf{y}_{1:n}|\mathbf{d}_{1:n}),$$

(e.g., Chaloner and Verdinelli (1995) and references therein). This is equivalent to the mutual information between the future observation $\mathbf{y}_{n+1}$ and the previous observations $\mathbf{y}_{1:n}$, conditional on the future designs $\mathbf{d}_{n+1}$ and previous designs $\mathbf{d}_{1:n}$. For normal linear models, this criterion is related to the classical c-optimal criterion (which maximises the precision of estimates of linear combinations of model parameters for linear models). Leipe et al. (2013) used mutual information to minimise prediction uncertainty in sequential systems biology experiments. Zidek et al. (2000) used maximum entropy to obtain designs that maximised information about expected responses for air quality monitoring sites.

Geostatistical design problems often use utilities that are functions of the prediction variance. For example, Diggle and Lophaven (2006) propose a Bayesian design criterion that chooses a set of sampling locations to enable efficient spatial prediction by minimising the expectation of the spatially averaged prediction variance (with respect to the marginal distribution of the data).

If one is interested in minimising the variance of the expected response, then one could use the utility function developed by Solonen et al. (2012) which places the next design point where the prior variance of the mean response is largest. The utility is calculated by bringing in the observations one-at-a-time and is given by:

$$U(\mathbf{d}, \mathbf{y}) = \prod_{k=1}^{K} (\sigma^2 + \text{Var}_{\boldsymbol{\theta}|\mathbf{y}_{1:(k-1)}}(m_k(\boldsymbol{\theta}))), \tag{2.4}$$

where $m_k(\boldsymbol{\theta}) = E(y_k|d_k, \boldsymbol{\theta})$ and K is the number of observations.
The expression $\text{Var}_{\boldsymbol{\theta}|\mathbf{y}_{1:(k-1)}}(m_k(\boldsymbol{\theta}))$ gives the variance of the mean response at d_k, given measurements $\mathbf{y}_{1:(k-1)}$ at points $\mathbf{d}_{1:(k-1)}$. The utility at d_k is evaluated using a weighted variance, where each simulated response is weighted based on the likelihood of previous simulated measurements, $p(\mathbf{y}_{1:(k-1)}|\mathbf{d}_{1:(k-1)}, \boldsymbol{\theta})$.

Solonen et al. (2012) advocate the use of this utility function to design for parameter estimation since it is easier to compute than information-based utility functions (equation (2.2)). Solonen et al.'s (2012) utility function assumes a constant variance. Ryan et al. (2014c) present a generalised version of this utility function which may be used when the error structure of a model has a non-constant variance.

2.4.4 Utilities for Several Design Objectives

Researchers often have several competing goals for an experiment, rather than one single goal, and so these competing design objectives can be incorporated into one or several utility functions. One approach to dealing with competing design objectives is to weight each design criterion and search for the design that optimises the weighted average of these criteria. This is known as a compound or weighted design problem (e.g., Dette (1990)). Clyde and Chaloner (1996) discuss compound design criteria and present an equivalence theorem for Bayesian constrained design problems. DasGupta et al. (1992) gave examples of compromise designs in which one is interested in finding a design that is highly efficient for several design problems.

Borth (1975) extends the mutual information utility proposed by Box and Hill (1967) so that fully Bayesian designs could be obtained for the dual goals of model discrimination and parameter estimation. This utility is known as "Total entropy". This dual design problem has been investigated in a number of classical design papers through use of compound criteria such as $D|T$- and $T|D-$optimality and hybrid DT-optimality (e.g., Atkinson (2008); Tommasi (2009); Waterhouse et al. (2009)), but is largely unexplored in the Bayesian design literature.

Chaloner and Verdinelli (1995) discuss several Bayesian utility functions that may be used for the dual purpose of maximising the expected value of the response and the expected information gain, and utilities which may be used to design for parameter estimation and prediction.

McGree et al. (2012c) considered compound utility functions in the context of Bayesian adaptive designs for dose-finding studies for the dual design objectives of estimating the maximum tolerated dose and addressing the safety of the study subjects. A number of different estimation utilities were used, and the utility functions only allowed doses to be available for selection if the 95th percentile of the posterior predictive probability of toxicity was less than some pre-specified tolerance level. Drovandi et al. (2013) developed a hybrid utility function for an adaptive dose-finding study to obtain robust estimates of the target stimulus-response curve in the presence of model and parameter uncertainty.

A number of studies have had the dual objectives of designing for parameter estimation or prediction accuracy and to minimise study costs (or inconvenience to study subjects). Stroud et al. (2001) used utility functions which designed for the precise estimation of parameters of interest, as well as minimising inconvenience to study subjects by penalising samples that were collected after a certain time period. Palmer and Müller (1998) searched for the optimal sampling times for stem cell collections in cancer patients, to minimise the expected loss function over the posterior predictive distribution for a new patient. Their utility function also included a penalty for failing to collect a certain target number of stem cells and a cost penalty for each sampling time scheduled.

2.5 Static Design Search Algorithms

Static design problems assume that the same design will be used throughout the experimental process, regardless of the incoming information that may be collected from the experiment. Static designs are useful when data are collected in a batch, according to a fixed protocol. Static designs are also useful for experiments in which data are not available until a considerable time after treatment allocation. A number of different algorithms have been used to solve Bayesian static design problems and they will be discussed below.

2.5.1 MCMC Algorithms

A number of stochastic algorithms have been proposed in the literature to approximate the maximisation and integration problem of equation (2.1) for static design problems. These include: prior simulation (Müller (1999)); smoothing of Monte Carlo simulations (Müller (1999)); MCMC simulation in an augmented probability model (Müller (1999)); and sequential Monte Carlo methods (Kück et al. (2006)).

Monte Carlo Integration

In many situations, one can simulate values of $(\boldsymbol{\theta}_i, \mathbf{y}_i)$ (for $i = 1, ..., M$) from $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ and the utility function can be estimated using these values. The integral is approximated

by using:

$$\hat{U}(\mathbf{d}) = \frac{1}{M}\sum_{i=1}^{M} U(\mathbf{d}, \boldsymbol{\theta}_i, \mathbf{y}_i). \tag{2.5}$$

The optimal design, $\mathbf{d}^* = \arg\max \hat{U}(\mathbf{d})$, can then be found by using a suitable maximisation method to search over the estimates, $\hat{U}(\mathbf{d})$ (see Müller (1999)). This approach has commonly been used in the literature (e.g., Wakefield (1994); Carlin et al. (1998); Palmer and Müller (1998)) and is useful when a discrete set of possible designs that are of low dimension are used.

Müller and Parmigiani (1995) use a similar approach to equation (2.5), in which stochastic optimisation is performed by fitting curves to the Monte Carlo samples. First, they simulate draws from $(\boldsymbol{\theta}, \mathbf{y})$ and evaluate the observed utilities. Then, a smooth curve is fitted through these simulated points, which serves as an estimate of the expected utility surface. The optimal design can then be found deterministically. Kuo et al. (1999) also used these curve fitting methods for solving design problems of low dimension.

Straightforward Monte Carlo integration over $(\boldsymbol{\theta}, \mathbf{y})$ for each design $\mathbf{d}$ may be computationally intensive for design problems involving a large number of design variables, since the design space grows far too rapidly with the number of design variables and thus the grid search over the design space becomes infeasible. Also, when a design variable corresponds to a data point, then a larger number of design variables means that more observations are involved, which implies a larger integral over $\mathbf{y}$, and thus a larger value of M is required to accurately estimate $U(\mathbf{d})$.

MCMC Simulation in an Augmented Probability Model

Alternatively, Clyde et al. (1996), Bielza et al. (1999) and Müller (1999) solved the optimal design problem by treating the expected utility as an unnormalised marginal probability density function. This was achieved by placing a joint distribution on $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ to form an augmented probability model $h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$, which is given by:

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}),$$

where it was assumed that $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ satisfies the appropriate conditions for $h(\cdot)$ to be positive and integrable over $(\mathbf{D}, \boldsymbol{\Theta}, \mathbf{Y})$. The probability distribution $h(\cdot)$ is defined such that the marginal distribution in $\mathbf{d}$ is proportional to the expected utility, i.e.,

$$\begin{aligned} h(\mathbf{d}) &\propto \int\int U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} \\ &= U(\mathbf{d}). \end{aligned}$$

It is assumed that the design space $\mathbf{D}$ is bounded and that the utility $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ is non-negative and bounded. One can then use a Metropolis-Hastings (MH) MCMC scheme to simulate from $h(\cdot)$ and select random draws from the design space that are proportional to the utility that is attached to the design. The MH MCMC algorithm focuses on sampling

designs in areas of high expected utility and discourages sampling in areas of low expected utility (see Müller (1999)). The sample of simulated $\mathbf{d}$ may be used to provide an estimate of $h(\mathbf{d})$ and the joint mode of $h(\mathbf{d})$, $\mathbf{d}^*$, corresponds to the optimal design.

We note that the joint mode of $h(\mathbf{d})$ needs to be found rather than the marginal modes for each element of $\mathbf{d}$ as the latter may be very different from the former. Cook et al. (2008) and Drovandi and Pettitt (2014) propose methods for searching for the multivariate mode using a non-parametric density estimate of the (annealed) expected utility surface based on the design samples obtained from the MCMC. However, for design problems that involve a large number of design points ($dim(\mathbf{d}) \geq 4$), the problem of finding the multivariate mode is more difficult than finding marginal modes and one may need to use dimension reduction techniques, such as those that Ryan et al. (2014c) propose. However, dimension reduction techniques may not always be appropriate and further research is needed into the problem of finding the multivariate mode for a large number of design variables.

In some design problems the range of values taken by the utility in the neighbourhood of its mode can be sufficiently small so that the Monte Carlo error can dominate this range. Then the mode is difficult to locate accurately. However, the problem can be mitigated by the fact that there is a neighbourhood of designs with near optimum utility and exact location of the mode is therefore not necessary.

Simulated Annealing-type Approach

Müller (1999) proposes an approach that is similar to simulated annealing (see Van Laarhoven and Aarts (1987)) in which the expected utility surfaces are replaced by a more peaked surface. This does not change the solution of the optimal design problem. The target function $h(\mathbf{d})$ is replaced with $h_J(\mathbf{d})$ where J is an integer, usually large (say 20 or higher). The joint augmented distribution to simulate from is:

$$h_J(\mathbf{d}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_J, \mathbf{y}_1, ..., \mathbf{y}_J) \propto \prod_{j=1}^{J} U(\mathbf{d}, \boldsymbol{\theta}_j, \mathbf{y}_j) p(\boldsymbol{\theta}_j, \mathbf{y}_j | \mathbf{d}). \tag{2.6}$$

For each $\mathbf{d}$, one simulates J experiments $(\boldsymbol{\theta}_j, \mathbf{y}_j)$, $j = 1, ..., J$, independently from $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ and considers the product of the calculated utilities. The product of the calculated utilities (rather than the sum) is used to ensure that $h_J(\mathbf{d}) \propto U^J(\mathbf{d})$.

This approach has been very popular in the literature (e.g., Bielza et al. (1999); Müller (1999); Stroud et al. (2001); Cook et al. (2008); Ryan et al. (2014c)) and uses similar ideas to simulated annealing (see Van Laarhoven and Aarts (1987)) where $T = 1/J$ may be interpreted as the "annealing temperature". As $T \to 0$, the original target function is replaced with a point mass at the mode (Müller (1999)). As J increases, the utility surface will become more peaked and simulations will cluster more tightly around the mode. However, increasing J obviously increases the number of required computations. An annealing schedule is not required, i.e., the same value of J may be used for all simulations. However, this is not efficient for high dimensional problems (see Amzal et al.

(2006)) and a "cooling" schedule may be required where J increases to $+\infty$. Müller et al. (2004) recommend that J should be gradually increased as the algorithm progresses so that the search will not become trapped in a local mode for situations where several modes exist. In Müller et al.'s (2004) approach, the algorithm initially explores the entire design space, but as the J value increases, the MCMC draws focus around one of the highest modes.

Whilst the algorithm presented by Müller (1999) has "theoretically appealing" properties (i.e., one can sample from the expected utility surface using a MH MCMC algorithm in which sampling is focused in areas of a high expected utility; and as $J \to \infty$, the expected utility is replaced with a point mass at the mode), it has been found to have slow convergence in practice, particularly for situations where there are a large number of design variables for which this algorithm becomes inefficient (Stroud et al. (2001); Amzal et al. (2006)). Use of this algorithm has therefore mostly been restricted to up to four design variables (e.g., Bielza et al. (1999); Müller (1999); Stroud et al. (2001); Cook et al. (2008)) and further research is required for searching for solutions to high dimensional design problems.

2.5.2 SMC Algorithms

Sequential Monte Carlo (SMC) algorithms, also known as "particle filters", use a population of particles to approximate a distribution and move through a smooth sequence of connected target distributions using resampling and diversification of particles until the final target distribution is reached (see Chopin (2002); Del Moral et al. (2006)). SMC combined with Markov and MCMC kernels provides a powerful and efficient computational approach for approximating target distributions. SMC has only been applied to static design problems in a few instances (see Amzal et al. (2006); Kück et al. (2006)).

SMC methods can be useful for sampling from target distributions that change over time. This also includes the target distribution $h_J(\mathbf{d}, \boldsymbol{\theta}_{1:J}, \mathbf{y}_{1:J})$ (Müller et al. (2004)) in which J increases over time. For nonlinear and high dimensional design problems, Amzal et al. (2006) extended the approach of Müller (1999) and Müller et al. (2004) through the use of particle methods, which are similar to particle filters (e.g., Doucet et al. (2001); Chopin (2002)) and population Monte Carlo simulations (e.g., Cappé et al. (2004)). This involves the simulation of N_p "parallel" Markov chains from the target distribution $h_J(\cdot)$, which are known as an "interacting particles system". At each iteration of the algorithm, an approximate weighted sample is generated from $h_J(\mathbf{d}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_J, \mathbf{y}_1, ..., \mathbf{y}_J)$ via importance sampling (Geweke (1989)) and a selection procedure, such as 'sampling with replacement' (e.g., Chopin (2002)), is then used to duplicate those particles that occur near the modes of the target distribution, whilst eliminating those that fall further away. An independent Markov step (Chopin (2002)) could also be added to the algorithm to avoid degeneracy problems and enrich the sample. Amzal et al. (2006) suggest that the proposal distribution for the design points should have a fairly large variance to enable the detection of other modes.

Amzal et al. (2006) also propose a "resampling-Markov algorithm" in which the importance sampling step is only implemented at initialisation of the algorithm. It is assumed that at time $t-1$, a sample that approximates $h_{J(t-1)}$ is available. If $J(t) > J(t-1)$, then additional values for each of the particles $\{\boldsymbol{\theta}_{j,k}^{(t-1)}, \mathbf{y}_{j,k}^{(t-1)}\}_{k=1}^{N_p}$, $j = J(t-1)+1, ..., J(t)$, are drawn from the posterior for $t-1$ and weights are computed that are proportional to the product of the utilities of the newly sampled values. The resampling and Markov steps are then implemented.

Kück et al. (2006) generalise the approach of Müller et al. (2004) to non integer annealing steps and used SMC samplers (similar to Del Moral et al. (2006); Johansen et al. (2006)) to search for designs which maximise the KLD between the prior predictive and posterior predictive distributions. In their application the model parameters $\boldsymbol{\theta}$ could be analytically integrated out of the expected utility which simplified the problem. A sequence of target distributions is generated artificially by incrementally "powering up" some measure of the utility for each particle. They define the following sequence of artificial target distributions:

$$h_{J(t),\nu_t}(\mathbf{d}, \boldsymbol{\theta}_{1:J(t)}, \mathbf{y}_{1:J(t)}) \propto \left(\prod_{j=1}^{J(t)-1} U(\mathbf{d}, \boldsymbol{\theta}_j, \mathbf{y}_j) p(\mathbf{y}_j|\mathbf{d}, \boldsymbol{\theta}_j) p(\boldsymbol{\theta}_j) \right) \times \left\{ U(\mathbf{d}, \boldsymbol{\theta}_{J(t)}, \mathbf{y}_{J(t)}) p(\mathbf{y}_{J(t)}|\mathbf{d}, \boldsymbol{\theta}_{J(t)}) p(\boldsymbol{\theta}_{J(t)}) \right\}^{\nu_t}.$$

The inverse annealing temperature, $J(t)$, is assumed to have an integer value. Kück et al. (2006) assumed that $J(t)$ could increase by at most one per iteration. Kück et al. (2006) also used the variable $\nu_t \in [0, 1]$ to enable smoother non-integer increases of the inverse annealing temperature. If $\nu_t = 1$, then the target density given in equation (2.6) is obtained and the dimension of the target distribution increases by $J(t) = J(t-1) + 1$, and ν_t is set back to zero. In all other instances, $J(t) = J(t-1)$. The choice of how to increase $J(t)$ is important, since large increments could result in degeneracy of the particles, and small increments are computationally inefficient. McGree et al. (2012a) propose to choose the increment to maintain a specific level of efficiency (based on the ESS) in the sample.

At time $t-1$, the particle set $\{\mathbf{d}_k^{(t-1)}, W_k^{(t-1)}\}_{k=1}^{N_p}$ provides an approximation for $h_{J(t-1)}$. A re-weight step is then implemented in the SMC algorithm via importance sampling to update the weighted particle set to approximate $h_{J(t)}$. Particles with a higher utility are given more weight than those with a lower utility. As J increases, the target distribution becomes more peaked around the mode. Resampling and mutation steps are also used to avoid denegeracy in the particle set.

Kück et al.'s (2006) approach was found to behave well when exploring multi-modal target distributions.

2.5.3 Other Stochastic Approximation Algorithms

Huan and Marzouk (2013) used simultaneous perturbation stochastic approximation (SPSA) (Spall (1998)) and Nelder-Mead nonlinear simplex (NMNS) (Nelder and Mead (1965)) algorithms to perform stochastic optimisation for nonlinear and computationally intensive models. SPSA is a stochastic approximation method that is similar in nature to a steepest-descent method that uses a finite difference estimate of the gradient. However, SPSA only uses two random perturbations to estimate the gradient, regardless of the dimension of the problem. Whilst the finite differences stochastic approximation (FDSA) algorithm only perturbs in one direction at a time, the SPSA algorithm perturbs in all directions at once. In SPSA, the error in the estimation of the gradient is "averaged out" over a large number of iterations (Spall (1998)) and the algorithm has a similar convergence rate to FDSA. SPSA has a global convergence property that relies on the existence of a non-negligable noise level in the objective function and the finite-difference-like perturbations (Maryak and Chin (2004)). However, high noise levels can cause slow convergence or can cause the algorithm to become stuck in local optima. SPSA is suitable for large-scale population models.

The NMNS algorithm has commonly been used for deterministic optimisation of nonlinear functions. It is a well-studied numerical method that is useful for problems in which gradients may be unknown. The NMNS algorithm is useful when dealing with noisy objective functions since it only requires a relative ordering of the function values, rather than the magnitudes of the differences (as when estimating gradients). NMNS is less sensitive than SPSA to the noise level, but can converge to non-stationary points. Huan and Marzouk (2013) found that the NMNS algorithm performed better than SPSA overall, in terms of the asymptotic distribution of the design variables and how quickly convergence was achieved.

Huan and Marzouk (2012) used the Robbins-Munro (RM) (Robbins and Monro (1951)) stochastic approximation, and compared it to a sample average approximation combined with the Broyden-Fletcher-Goldfarb-Shanno method (SAA-BFGS) to solve the optimal design problem for partial differential equations. The RM algorithm is one of the oldest stochastic approximation methods. It uses an iterative update that is similar to steepest descent, but uses stochastic gradient information. Sampling average approximation (SAA) algorithms reduce a stochastic optimisation problem to a deterministic one. For instance, in the optimal experimental design framework, we may define the problem to be solved as:

$$\mathbf{d}^* = \arg\max_{\mathbf{d}\in\mathbf{D}}\{U(\mathbf{d})\} = E_W[\hat{U}(\mathbf{d}, W)],$$

where $\mathbf{d}$ is the design variable, W is the "noise" random variable, and $\hat{U}(\mathbf{d}, W)$ is an unbiased estimate of the objective function, $U(\mathbf{d})$ (e.g., KLD between the prior and posterior distributions). SAA approximates this optimisation problem using

$$\hat{\mathbf{d}}_s = \arg\max_{\mathbf{d}\in\mathbf{D}}\{\hat{U}_M(\mathbf{d}, w_s) \equiv \frac{1}{M}\sum_{i=1}^{M}\hat{U}(\mathbf{d}, w_i)\},$$

where $\hat{\mathbf{d}}_s$ and $\hat{U}_M(\mathbf{d}, w_s)$ are the optimal design and utility function values under a particular set of M realisations of W, where $w_s \equiv \{w_i\}_{i=1}^M$. The same set of realisation of W is used for different values of $\mathbf{d}$ throughout the optimisation process, which makes the maximisation problem deterministic. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (Nocedal and Wright (2006)), which is a deterministic quasi-Newton method, was used to find $\hat{\mathbf{d}}_s$ as an approximation to $\mathbf{d}^*$.

Huan and Marzouk (2012) used infinitesimal perturbation analysis (Ho and Cao (1983)) to construct an unbiased estimator of the gradient of the KLD for use in the RM algorithm. A polynomial chaos approximation of the forward model was also used to speed up computation of the utility function and gradient evaluations. Huan and Marzouk (2012) found that, although SAA-BFGS generally required fewer iterations, each iteration had a longer run time than a step of RM. As the evaluation of the utility function becomes more expensive, RM may be the more suitable of the two methods. RM was also found to outperform SAA-BFGS in terms of the size of the mean square error (between the "true" optimal value of the KLD and the value of the KLD for the current iteration), for a given computational effort.

2.6 Sequential Design Search Algorithms

Decisions are often made in stages, with additional data being observed between the decisions. For example, in dose-finding trials, dose allocation decisions are often made after previous cohorts have been administered the treatment so that future cohorts may be given doses that are closer to the maximum tolerated dose. Whitehead and Brunier (1995) and Whitehead and Williamson (1998) implement a Bayesian m-step look-ahead procedure to find the optimal treatment dose to administer to the next m patients in a dose-finding study. Sequential design problems are those that involve an alternating sequence of decisions and observations. The Bayesian paradigm is extremely useful for sequential design problems since the posterior can be used as the prior distribution for the next experiment.

2.6.1 Backwards Induction

Although many approaches to solving sequential design problems use a myopic approach, which involves looking ahead only to the next observation (e.g., Cavagnaro et al. (2010); Drovandi et al. (2014); McGree et al. (2012b)), in general, this is not optimal, and one should instead look ahead to all future observations in the experiment (Borth (1975)), as well as the decisions that might be made at each future observation. To achieve this, the computationally intensive *backward induction* method should be used (see, for example, DeGroot (1970); Berger (1985); Bernardo and Smith (2000) for a description) which considers all future observations. Backward induction is also known as stochastic dynamic programming (e.g., Ross (1983)).

Early work in this area was restricted to simple model settings, such as one-sided tests of a univariate parameter (Berry and Ho (1988)), and binary outcome settings (Lewis and Berry (1994)). These approaches typically used only two or three backwards steps

(interim looks at the data). Carlin et al. (1998) extend these approaches by including a forward sampling algorithm that can be used to find the optimal stopping boundaries in clinical trials and eases the computational burdens associated with backward induction. However, Carlin et al. (1998) used a univariate normal likelihood, assumed that the standard deviations were known at each step, and considered a maximum of 4 backwards steps.

Brockwell and Kadane (2003) proposed a gridding method which approximates the expected loss function (utility function) at each decision time, and consists of a function of certain summary statistics (low dimensional) of the posterior distribution of the parameter of interest. Their approach is similar to that of Berry et al. (2000). Brockwell and Kadane (2003) use a one-step-ahead forward simulation procedure to evaluate the expected utilities and focus on problems related to parameter estimation. Müller et al. (2006) also use a similar approach to Brockwell and Kadane (2003) which involves forward simulation to approximate the utility functions and constrain the action space to circumvent the problem of an increasing number of possible trajectories in the backward induction steps. Rossell et al. (2007) extend the approaches of Carlin et al. (1998), Brockwell and Kadane (2003), and Müller et al. (2006), in which they compute a summary statistic when new data are observed and use decision boundaries that partition the sample space. Once the summary statistic falls in the stopping region, the experiment is terminated. Thus the sequential problem is reduced to the problem of finding optimal stopping boundaries, and the choice of these boundaries accounts for all future data. Rossell and Müller (2013) extend these ideas to high dimensional data by assuming that the data are suitably pre-processed.

2.6.2 MCMC Algorithms

McGree et al. (2012c) used MCMC methods (MH algorithms) to sample from the posterior distribution to find adaptive designs for a dose-finding study. Bayesian compound utility functions were used to find the dose for the next subject for the dual purposes of estimating the maximum tolerated dose (MTD) and addressing safety issues of toxicity. To estimate the utility functions, importance sampling was used in which the posterior distribution of the parameters (using the observations up to the $i-1$th subject) $p(\boldsymbol{\theta}|\mathbf{y}_{(1:i-1)})$ was used as the importance distribution, and the target distribution was $p(\boldsymbol{\theta}|\mathbf{y}_{(1:i)})$, where $\mathbf{y}_i$ is the new data point given by dose D. McGree et al.'s (2012c) algorithm involved a form of self-tuning in that the proposal distribution for the model parameters $\boldsymbol{\theta}$ was based on a bivariate normal distribution in which the mean and variance were obtained from a maximum likelihood fit to the current data. Each time a new dose was selected, the proposal distribution was updated.

2.6.3 SMC Algorithms

SMC provides a natural framework for sequential design problems and has been used for parameter estimation design problems (e.g., Drovandi et al. (2013)), and model discrimination design problems (see Cavagnaro et al. (2010); Drovandi et al. (2014)). Its design

applications are diverse and include computer experiments (e.g., Loeppky et al. (2010)), astrophysics (e.g., Loredo (2004)), cognitive science (e.g., Cavagnaro et al. (2010)), neurophysiology experiments (e.g., Lewi et al. (2009)), clinical trials (e.g., Liu et al. (2009)) and bioassays (e.g., Tian and Wang (2009)).

SMC algorithms have commonly been used to design for model discrimination for Bayesian sequential design problems (e.g., Cavagnaro et al. (2010); Drovandi et al. (2014); McGree et al. (2012b)). Designs that efficiently and precisely estimate model parameters usually perform poorly on model discrimination problems (e.g., Atkinson (2008)).

Cavagnaro et al. (2010) use a similar approach to Amzal et al. (2006) in which an SMC algorithm was implemented to design optimally for model discrimination in the context of memory retention models. A simulated annealing effect (Müller (1999)) was used in which the utility function was incrementally "powered up". Cavagnaro et al.'s (2010) SMC algorithm designs for experiments one-observation-at-a-time, using the posterior distribution that is based on all of the data that has been observed thus far. Whilst these myopic approaches are sub-optimal, they are necessary in many applications of Bayesian design of experiments due to computational complexity of the backwards induction algorithm (Section 2.6.1).

Drovandi et al. (2014) present an SMC algorithm to sequentially design experiments one-at-a-time in the presence of model uncertainty for discrete data. McGree et al. (2012b) extended this approach for continuous data. In these works, an SMC algorithm is run in parallel for each of the competing models and the results are combined to compute the utility function in the presence of model uncertainty. This algorithm avoids between-model or cross dimensional proposals. The SMC algorithm produces an approximation to the evidence (the marginal likelihood of the data given a particular model) as a by-product (Del Moral et al. (2006)), which is used to compute the posterior model probabilities and to estimate the utility function. This avoids the need to use computationally intensive numerical integration techniques, such as quadrature (e.g., Cavagnaro et al. (2010)) to obtain an estimate of the evidence. Once the posterior model probabilities are computed, model discrimination utility functions, that are derived from information theory, such as the entropy of model probabilities (Box and Hill (1967); Borth (1975)) can be evaluated. The design d that is chosen is the one that maximises the mutual information between the model indicator, m, and the predicted observation (Cavagnaro et al. (2010)). Little problem specific tuning is required for this algorithm and it is much less computationally intensive than approaches that rely on MCMC for posterior simulation in sequential design contexts (e.g., McGree et al. (2012c), Section 2.6.2).

In both Drovandi et al. (2014) and McGree et al.'s (2012b) work, only a discrete design space was considered and no optimisation algorithm was implemented. To reduce the computational requirements, the utility was evaluated for all possible choices of design, and the design which maximised the utility was chosen. For high dimensional design problems or those with continuous support, optimisation routines such as the exchange algorithm (Meyer and Nachtsheim (1995)) or simulated annealing (Corana et al. (1987))

may be required. Alternatively, one could incorporate the simulation-based algorithms of Müller (1999) or Amzal et al. (2006) to search over the design space.

2.7 Applications

We will now highlight some of the key areas that Bayesian experimental design is being applied to. Please note that this is not a comprehensive review on the applications of Bayesian experimental design, but rather an overview of some of the key papers in the literature.

2.7.1 Clinical Trial Design

There is a wealth of literature on Bayesian designs for early stage clinical trial studies (phase I and II clinical trials where dosage is investigated), with many practical developments in both sequential and static frameworks. The requirement of a decision-theoretic approach for the design of clinical trials was recognised as early as Anscombe (1968). Clinical trial design typically involves making decisions prior to the commencement of the experiment in relation to the drug dosage (e.g., dose level, timing of doses, number of doses, number of subjects to assign to each dose level), and for pharmacokinetic (what the subjects' body does to the treatment) and pharmacodynamic (what the treatment does to the subjects' body) studies, sampling times (e.g., number of samples to take, timing of the samples, assignment of subjects to sampling schedules).

Spiegelhalter (2004) and Berry (2006) provide a general overview on how Bayesian methods can be used for inference and experimental design for clinical trial design. Berry (1993), Spiegelhalter et al. (1994), Kadane (1996), and Stangl and Berry (1998) discuss a number of important issues present in the use of Bayesian methods for the design and analysis of clinical trials, such as: ethics, prior elicitation, randomisation, treatment allocation, utilities, and decision making. However, the use of fully Bayesian designs for clinical trial design remains mostly theoretical and their use in practice is still uncommon. This is most likely due to the difficulties associated with prior elicitation for complex models and selection of a utility function, as well as the computational difficulties of optimising the expected utility function.

Decisions are often made sequentially in clinical trials as information is gathered on the experimental process from previous study subjects and enables the investigators to modify the experiment according to the accumulated information. Modifications to sequential clinical trial designs include adaptively assigning study subjects to treatments that have a higher performance or that will give more information about the experimental aims; adding or deleting treatment arms; termination of the trial; and incorporation of more study subjects if the experimental aims have not been satisfied. Bornkamp et al. (2011), Müller et al. (2006), Wathen and Thall (2008), and Dragalin et al. (2010) present Bayesian sequential designs for clinical trials.

Dose-finding studies are concerned with determining the effect of different doses of the treatment on the response of interest and are required to ensure that marketed drug

doses are safe and efficacious. One of the earliest works on Bayesian adaptive design for dose-finding studies is that by Whitehead and Brunier (1995), in which they obtain priors through elicited data and select treatment doses based on the gain in statistical information about an estimate.

Bornkamp et al. (2011) determine adaptive designs in a dose-finding study so that the minimum effective dose (MED), i.e., the smallest dose that achieves a clinically beneficial response over the placebo response, can be precisely estimated (using the approaches described in Bornkamp et al. (2007)). The dose-response relationship of a treatment is often unknown prior to the study. To account for model uncertainty, Bornkamp et al. (2011) average the design criterion (the posterior variance of the MED), conditional on model m, with respect to the model probabilities (see also, Dette et al. (2008)). This produces designs that are robust to model uncertainty. Bornkamp et al. (2011) also use a Bayesian shrinkage approach to stabilise the parameter estimates during the sequential updates of the parameter estimates and model probability.

Müller et al. (2006) use a grid-based backwards induction approach to make decisions about adaptive dose allocation, optimal stopping of a trial and the optimal decision upon stopping to enable optimal learning about the dose-response curve. Wathen and Thall (2008) describe an approach to find a group sequential design that maintains a targeted false-positive rate and power, under a wide range of true event time distributions for right-censored data in a phase III clinical trial. At each interim analysis, Wathen and Thall's (2008) procedure adaptively chooses the most likely model (based on the posterior probability) for the hazard function, using Bayesian model selection, and then they apply the decision bounds that are optimal for the chosen model. Their focus is on two-sided tests in two-arm trials. Dragalin et al. (2010) conduct an extensive simulation study that compares five different adaptive dose-finding designs. These designs differed in the number of doses, the number of interim analyses, and the number of patients allocated to each design, and were derived under different experimental objectives.

Stroud et al. (2001) use the MH MCMC algorithm of Müller (1999) to determine the optimal blood sampling times for the next patient to precisely estimate pharmacokinetic parameters of interest (subject to a cost penalty) for the anticancer agent, paclitaxel. The priors were obtained by fitting nonlinear mixed effects models to existing data. Ryan et al. (2014b) present fully Bayesian static designs for a horse population pharmacokinetic study. The design problem was to determine the optimal urine sampling times, as well as the number of subjects and samples per subject to obtain precise posterior distributions of the population parameters (subject to a cost constraint). These designs were also obtained using an adaption of the MH MCMC algorithm of Müller (1999).

Other recent examples of Bayesian clinical trial design include Christen et al. (2004), Dragalin et al. (2007), Miller et al. (2007), Ding et al. (2008), Drovandi et al. (2013), and McGree et al. (2012c).

2.7.2 Cognitive Science

Optimal experimental design methods are also commonly used in the field of cognitive science. For example, Kujala and Lukka (2006) and Lesmes et al. (2006) use Bayesian sequential designs to estimate psychometric functions using utility functions based on maximum entropy. Lewi et al. (2009) present a sequential design framework that searches for the optimal design for a neurophysiology experiment that maximises the mutual information between the prior and posterior distribution for a generalised linear model. To facilitate estimation of the mutual information, a Gaussian approximation to the posterior was used. Myung and Pitt (2009) search for optimal static designs for model discrimination for memory retention and categorisation examples, using the approach of Amzal et al. (2006). Cavagnaro et al. (2010) use a similar approach to Myung and Pitt (2009), but instead search for sequential designs for model discrimination for a memory retention example. Kim et al. (2013) extend the approach of Cavagnaro et al. (2010) to find sequential designs for a population study of visual perception. Zhang and Lee (2010) find sequential designs to discriminate amongst competing models for a two arm bandit problem for human choice behaviour.

2.7.3 Natural Sciences

Loredo (2004) found sequential Bayesian designs for astrophysics experiments to detect extrasolar planets, using the maximum expected Shannon information gain on the posterior parameter estimates for the utility function. Huan and Marzouk (2013) use polynomial chaos approximations and nested Monte Carlo integration (Ryan (2003)) to estimate the KLD between the prior and posterior distributions to find static designs which enable inference about parameters for chemical kinetic models for combustion. Solonen et al. (2012) derive optimal static designs for an exothermic example. Cook et al. (2008) used Bayesian simulation-based strategies similar to Müller (1999) to determine observation times for botanical epidemic experiments that were governed by nonlinear stochastic processes.

2.8 Directions for Future Research

We believe the future of Bayesian experimental design lies in: (1) developing and implementing fast methods for approximating the posterior distribution for use in Bayesian utility functions, and fast computation of the Bayesian utility functions, as these are the most computationally intensive components of Bayesian experimental design; and (2) finding solutions to complex Bayesian experimental design problems, such as problems in which the likelihood is intractable or computationally prohibitive to calculate, or problems with a large number of design points.

2.8.1 Fast Algorithms for Bayesian Experimental Design

In Table 2.1 we provide a summary of the methods which have previously been used to approximate the posteriors for Bayesian utility functions, along with the search algorithms in which they are embedded.

Search Algorithm framework	Method for approx. posterior	Example(s)
Static designs		
MCMC	Laplace approximation	Ryan et al. (2014a)
MCMC	Importance sampling	Cook et al. (2008); Ryan et al. (2014a,c)
MCMC	ABC	Drovandi and Pettitt (2013); Hainy et al. (2013)
MCMC	MCMC	Clyde et al. (1996)
Monte Carlo	MCMC	Han and Chaloner (2004)
SMC	Importance sampling	Amzal et al. (2006)
SPSA and NMNS	Polynomial chaos approximations and nested Monte Carlo integration	Huan and Marzouk (2013)
RM stochastic approximation	Nested Monte Carlo integration	Huan and Marzouk (2012)
SAA-BFGS	Nested Monte Carlo integration	Huan and Marzouk (2012)
Sequential designs		
Discrete search	Laplace approximation	Lewi et al. (2009)
SMC	Numerical quadrature	Cavagnaro et al. (2010)
Discrete search	SMC / importance sampling	Drovandi et al. (2013)
MCMC	Importance sampling	Stroud et al. (2001); McGree et al. (2012c)
Monte Carlo	MCMC	Wakefield (1994); Palmer and Müller (1998)

Table 2.1: Summary of methods used to approximate the posterior distributions for Bayesian utility function estimation and for optimisation over $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$.

MCMC and importance sampling have been found to be computationally intensive to perform at each iteration of the optimisation algorithm that searches over the space $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$, due to the large number of samples that are required to ensure that the Bayesian utility is well estimated. In particular, importance sampling from the prior performs poorly when large amounts of data are involved due to a low ESS (Ryan et al. (2014a)). Adaptive importance sampling (e.g., Kinas (1996); Pennanen and Koivu (2006)) may provide a faster method for approximating the posterior distributions, but is yet to be explored for Bayesian experimental design.

Laplace approximations and numerical quadrature have been found to be fast alternatives for approximating the posterior distribution in Bayesian design, and can be used when large amounts of data are involved, but rely on the assumption that the posterior distribution follows a multivariate normal distribution and also suffer from the curse of dimensionality. INLA can also provide a fast method for approximating the posterior distribution, but has not been used for Bayesian experimental design. VB methods are a fast method for facilitating approximate inference for intractable posterior distributions, but are yet to be used in a Bayesian experimental design context.

Drovandi and Pettitt (2013) and Hainy et al. (2013) have explored the use of ABC rejection (see Beaumont et al. (2002)) within an MCMC framework to approximate the posterior distributions for Bayesian utility functions for design problems in which the likelihood function is intractable. Further use of ABC methods for posterior distribution approximation should be explored in the experimental design context.

A few studies have investigated the use of SMC for approximating the necessary quantities for Bayesian utility functions (e.g., Drovandi et al. (2013)), but its use has been limited. Future studies should focus on extending previous approaches to allow for more complicated design problems. SMC with a Liu West filter (Liu and West (2001)) could offer a fast method for posterior approximation for Bayesian design problems.

Computational burden is a major obstacle in all Bayesian design problems for complex models and must be overcome so that designs can be obtained efficiently and in real time, and to broaden the applicability of Bayesian design methodology by making it more accessible to practitioners, scientists and industry. This may be achieved through algorithmic developments and the exploitation of current parallel computing technology (such as graphics processing units or GPUs). Indeed, new parallel architectures are becoming increasingly available to individual researchers, and will have a significant impact on Bayesian experimental design. In order to take advantage of this increased power, computational problems and approaches should be adapted from the current serial processing paradigm to one that optimises algorithms for parallel processing. To our knowledge, there is no published, peer reviewed research on the use of GPUs in the derivation of a Bayesian experimental design.

2.8.2 Finding Optimal Designs for Complex Models

The future of Bayesian experimental design also lies in solving complex or nonstandard problems, such as problems in which the likelihood is intractable or computationally prohibitive to evaluate, problems where the observed data likelihood cannot be evaluated analytically, or problems with a large number of design points. Whilst sophisticated inference techniques are available for Bayesian data analysis for complex data models, corresponding methodology for deriving Bayesian experimental designs is severely lacking, and it is important that the methods for inference are complemented with appropriate experimental design methodologies that enable more informative data to be collected in a more timely manner. Use of parallel computing technology may be required to ease the computational burden of finding optimal Bayesian experimental designs for complex models (such as mixed effects models).

Fully Bayesian experimental designs for nonlinear mixed effects models are largely unexplored. Most of the current work has focused on evaluating Bayesian utility functions for a fixed set of discrete designs (e.g., Han and Chaloner (2004); Palmer and Müller (1998)) and selecting the design that produces the highest utility value (i.e., no search over a continuous design space is performed). Ryan et al. (2014b) extend this by searching over a continuous design space to determine (near) optimal sampling times for a horse population pharmacokinetic study. Kim et al. (2013) find optimal sequential designs for population studies. Further work on using SMC algorithms (Chopin (2002)) to search for optimal designs for mixed effects models in the presence of model uncertainty is currently being conducted, so that solutions to real-world design problems can be found. The main

difficulty in finding solutions to experimental design problems in which the data is modelled by mixed effects models is that the observed data likelihood is unavailable in closed form for all but the simplest examples.

2.8.3 Finding Optimal Designs for a Large Number of Design Variables

Better search algorithms are also required to find static designs. Many of the search algorithms for obtaining optimal designs (e.g., Müller (1999); Amzal et al. (2006)) are restricted to a small number of design variables (≤ 4), as these algorithms are computationally prohibitive for a large number of design variables (e.g., Bielza et al. (1999); Müller (1999); Stroud et al. (2001); Cook et al. (2008)). MCMC algorithms are good at estimating the marginal distribution of random variables, but experimental design requires the joint distribution, and in particular the joint mode of the design variables, which is quite difficult to find and estimate.

Ryan et al. (2014c) propose the use of lower dimensional parameterisations to enable near optimal designs to be found for problems that require a large number of design points. The lower dimensional parameterisations consist of a few design variables, which are optimised, and are then input into various functions to generate multiple design points. This was found to have substantial computational savings, and it was much easier to obtain the multivariate mode for a few design variables than for a large number of design variables. However, designs found using this method are not optimal but *near* optimal, which is a compromise of the computational savings achieved. The approach is only useful for design variables (e.g., sampling times/locations) that require multiple measures to be taken at specific points that are separated from one another in the design space. This approach does not overcome the problem of having a large number of different types of design variables (e.g., temperatures, pressures), and further research needs to be conducted for solving this design problem.

2.9 Conclusion

Bayesian experimental design is a fast growing area of research with many exciting recent developments. The Bayesian approach to experimental design offers many advantages over frequentist approaches, the most notable of which is the ability to optimise design criteria that are functions of the posterior distribution and can easily be tailored to the experimenters' design objectives. Bayesian design criteria are optimised often with the assumption that Bayesian inference will be performed on the data that is obtained from the experimental design. Bayesian frameworks also provide a formal approach for incorporating parameter uncertainties and prior information into the design process via prior distributions, and provide a unified approach for joining these quantities with the model and design criterion. Another advantage of using a Bayesian design criterion is that a single design point can be used, and the prior distribution is updated by the single observation in a sequential manner. The prior information is not "thrown away" in fully Bayesian experimental design, as it is in pseudo-Bayesian design.

Whilst several review papers on Bayesian experimental design have been written, there is a lack of recent Bayesian experimental design papers that reflect the computational advancements that have occurred in recent times. In this article we have reviewed the computational methods that have been used to approximate the posterior distribution for Bayesian utility functions, along with methods for calculating the Bayesian utility functions (once the posterior has been approximated) and the search algorithms that have been used for finding the optimal designs. We have also highlighted some numerical methods and stochastic algorithms that have previously been used to perform Bayesian inference, but have not been used in the design context, and may provide fast alternatives for finding Bayesian designs.

It is our opinion that the future of Bayesian experimental design lies in the development and implementation of rapid methods for approximating the Bayesian utility functions, since this is the most computationally intensive component of the Bayesian experimental design process. We also believe that the future of Bayesian experimental design lies in finding solutions to complex or nonstandard design problems, such as problems in which the likelihood is intractable or computationally prohibitive to evaluate, problems where the observed data likelihood cannot be evaluated analytically, or problems with a large number of design points or design variables. Solutions to these difficult problems can only be achieved through algorithmic developments and the exploitation of current parallel computing technology.

Acknowledgments

E.G. Ryan was supported by an APA(I) Scholarship which came from an ARC Linkage Grant with Roche Palo Alto (LP0991602). The work of A.N. Pettitt was supported by an ARC Discovery Project (DP110100159), and the work of J.M. McGree was supported by an ARC Discovery Project (DP120100269).

Statement of Authorship for Chapter 3

This chapter has been written as a journal article. The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
- They agree to the use of the publication in the students thesis and its publication on the QUT ePrints database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for the associated publication is: **Ryan, E.G.**, Drovandi, C.C., Thompson, M.H., and Pettitt, A.N. (2014). Towards Bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics and Data Analysis*, 70: 45-60.

Contributor	Statement of contribution
Elizabeth Ryan	Wrote all Matlab code required for the manuscript, performed all computations in the manuscript, interpreted and reported the results, constructed all figures presented in the manuscript, wrote the manuscript, and acted as the corresponding author

Signature and Date:

Christopher Drovandi	Assissted in the writing of the Matlab code, directed the research and proofread the manuscript.
Helen Thompson	Directed the research and proofread the manuscript.
Tony Pettitt	Directed the research and proofread the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all co-authors confirming their certifying authorship.

Name	Signature	Date

Chapter 3

Towards Bayesian Experimental Design for Nonlinear Models that Require a Large Number of Sampling Times

ABSTRACT

The use of Bayesian methodologies for solving optimal experimental design problems has increased. Many of these methods have been found to be computationally intensive for design problems that require a large number of design points. A simulation-based approach that can be used to solve optimal design problems in which one is interested in finding a large number of (near) optimal design points for a small number of design variables is presented. The approach involves the use of lower dimensional parameterisations that consist of a few design variables, which generate multiple design points. Using this approach, one simply has to search over a few design variables, rather than searching over a large number of optimal design points, thus providing substantial computational savings. The methodologies are demonstrated on four applications, including the selection of sampling times for pharmacokinetic and heat transfer studies, and involve nonlinear models. Several Bayesian design criteria are also compared and contrasted, as well as several different lower dimensional parameterisation schemes for generating the many design points.

KEYWORDS: Bayesian optimal design; Sampling strategies; Robust design; Markov chain Monte Carlo; Stochastic optimisation.

3.1 Introduction

Optimal experimental design provides rules for the allocation of resources for studies which require data collection but where there is variability present, whether it is under the full control of the experimenter or not. Experimental designs are concerned with the incorporation of features into studies to control systematic error (bias), reduce random variations, and increase precision. Experimental design problems are commonly viewed as optimisation problems. Optimal experimental designs may be used to achieve the experimental goals more rapidly and hence reduce experimental costs.

Experimental design has been widely developed within the classical framework, in both theory and practice (e.g., Atkinson and Donev (1992)). In the classical framework, optimal experimental designs are commonly derived using optimality criteria that are based on the Fisher information matrix (e.g., Fedorov (1972); Pukelsheim and Torsney (1991); Atkinson and Donev (1992)).

Classical experimental design is well suited to linear or linearised models. However, for nonlinear models, designs are locally dependent on the values which are chosen for the model parameters. Since interest is often focused on the design of experiments that can provide accurate parameter estimates, this means that selection of the parameters from which to construct the design is of critical importance and the use of unsuitable parameter values may result in suboptimal designs. In the classical framework, for nonlinear models, only locally optimal designs can be obtained. To overcome the dependence of the design on the initial estimates of the parameters, several studies have incorporated probability distributions on the parameters and average local design criteria over the distribution so that the designs obtained may be robust to the initial choice of the parameter values (e.g., D'Argenio (1990); Duffull et al. (2005); Ogungbenro and Aarons (2007); Duffull et al. (2012)). However, these so-called "pseudo-Bayesian" design criteria are more computationally intensive than the classical design criteria.

Here we consider a Bayesian approach, and the use of Bayesian methodologies for optimal experimental design has increased in the last few years (e.g., Müller (1999); Stroud et al. (2001); Amzal et al. (2006); Müller et al. (2006); Cook et al. (2008)). When constructing Bayesian optimal designs, the maximum expected utility principle is employed (see DeGroot (1970)), in which the preferences of the decision maker are assumed to be encoded by a utility function, $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$. The utility function describes the worth of choosing the design $\mathbf{d}$ from the design space $\mathbf{D}$, yielding data $\mathbf{y}$ from a sample space $\mathbf{Y}$, given model parameters (and latent variables) $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The form of the utility function is specific to the application and should incorporate the experimental aims (see Lindley (1972); Chaloner and Verdinelli (1995)). A probabilistic model, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$, is also required for all relevant random variables and future data. The probability model is decomposed into a prior distribution $p(\boldsymbol{\theta})$ and a sampling distribution $p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta})$.

Lindley (1972) suggests that the choice of a design should be regarded as a decision problem and that the design $\mathbf{d}$ which maximises the expected utility should be selected. Historical data may be incorporated into the model by conditioning the decision process on the available information. The Bayesian framework seeks to determine the optimal design, $\mathbf{d}^*$, that maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$ with respect to the unknown future observations $\mathbf{y}$ and model parameters $\boldsymbol{\theta}$:

$$\begin{aligned} \mathbf{d}^* &= \arg\max_{\mathbf{d}\in\mathbf{D}} \; E[U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})] \\ &= \arg\max_{\mathbf{d}\in\mathbf{D}} \; \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y}. \end{aligned} \tag{3.1}$$

In this article we assume that $\mathbf{d}^*$ exists.

There are several difficulties associated with maximising the expected utility surfaces. First, the decision space is often quite intricate in practical problems. Second, equation (3.1) does not usually have a closed form solution and requires multiple integrations and a maximisation (often) over a large decision space to obtain the optimum, $\mathbf{d}^*$. Even if the decision space is small and easily parameterised, the utility function may be difficult

to integrate. Unless the likelihood and prior are specifically chosen to enable analytic evaluation of the integration problem, the maximisation and integration problem requires numerical approximation or stochastic solution methods. In this article we will focus on stochastic solution methods.

A range of stochastic algorithms have been proposed in the literature to approximate the maximisation and integration problem. These include: prior simulation (Müller (1999)); smoothing of Monte Carlo simulations (Müller (1999)); gridding methods which involve numerical quadrature or Laplace approximations (Brockwell and Kadane (2003)); Markov chain Monte Carlo (MCMC) simulation in an augmented probability model (Müller (1999)); and sequential Monte Carlo methods (Amzal et al. (2006)). Most of these simulation methods are based on the assumption that the integral(s) (equation (3.1)) may be evaluated by Monte Carlo simulations with relative ease. In the majority of situations, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ is available for efficient random variable generation and the utility function can be evaluated point-wise using the simulated $(\boldsymbol{\theta}_i, \mathbf{y}_i)$ for $i = 1, ..., M$. The integral may then be approximated by using:

$$\hat{U}(\mathbf{d}) = \frac{1}{M} \sum_{i=1}^{M} U(\mathbf{d}, \boldsymbol{\theta}_i, \mathbf{y}_i). \tag{3.2}$$

One can then use $\hat{U}(\mathbf{d})$ to find the optimal design, $\mathbf{d}^* = \arg\max \hat{U}(\mathbf{d})$, by using a suitable maximisation method (see Müller (1999)). However, straightforward Monte Carlo integration over $(\boldsymbol{\theta}, \mathbf{y})$ for each design $\mathbf{d}$ can be computationally intensive for high dimensional design problems since a large value of M is required to obtain reasonable accuracy of the estimate of $U(\mathbf{d})$.

Clyde et al. (1996), Bielza et al. (1999) and Müller (1999) instead treated the expected utility described in equation (3.1) as an unnormalised marginal probability density function. This was achieved by placing a joint distribution on $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ to form an augmented probability model $h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$, which is given by:

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}) p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}),$$

assuming that $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ satisfies the appropriate conditions for $h(\cdot)$ to be positive and integrable over $(\mathbf{D}, \boldsymbol{\Theta}, \mathbf{Y})$. The resulting probability distribution $h(\cdot)$ is defined such that the marginal distribution of $\mathbf{d}$ is proportional to the expected utility, i.e.,

$$h(\mathbf{d}) \propto \int \int U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y} = U(\mathbf{d}).$$

Usually it is assumed that the design space $\mathbf{D}$ is bounded and that the utility $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ is non-negative and bounded. One can then simulate from $h(\cdot)$ using, say, a Metropolis-Hastings MCMC scheme to solve the optimal design problem by selecting random draws from the design space that are proportional to the utility that is attached to the design. The MCMC simulation focuses on sampling in areas of high expected utility and discourages sampling in areas of low expected utility (see Müller (1999)). The sample

of simulated $\mathbf{d}$ may be used to provide an estimate of $h(\mathbf{d})$ and the mode of $h(\mathbf{d})$, $\mathbf{d}^*$, corresponds to the optimal design. We note that the joint mode of $h(\mathbf{d})$ needs to be found rather than the marginal modes for each element of $\mathbf{d}$ as the latter may be very different from the former. The MCMC samples also enable one to investigate the sensitivity of the design problem with relative ease, since a single optimum may not exist or there may be many designs with similar efficiency.

However, the shape of the expected utility surface, and thus $h(\mathbf{d})$, can be flat around its mode and prohibitively large simulation sample sizes may be required to estimate the mode. To overcome these problems, a sample is instead simulated from $h_J(\mathbf{d})$ where J is an integer, usually large (say 20 or higher). The joint augmented distribution to simulate from is now:

$$h_J(\mathbf{d}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_J, \mathbf{y}_1, ..., \mathbf{y}_J) \propto \prod_{j=1}^{J} U(\mathbf{d}, \boldsymbol{\theta}_j, \mathbf{y}_j) p(\boldsymbol{\theta}_j, \mathbf{y}_j | \mathbf{d}).$$

That is, for each $\mathbf{d}$, one simulates J experiments $(\boldsymbol{\theta}_j, \mathbf{y}_j), j = 1, ..., J$, independently from $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ and considers the product of the calculated utilities. The product of the calculated utilities (rather than the sum) is used to ensure that $h_J(\mathbf{d}) \propto U^J(\mathbf{d})$.

This relies on the same idea as in simulated annealing (see Van Laarhoven and Aarts (1987)) where $T = 1/J$ may be interpreted as the 'annealing temperature'. As $T \to 0$, the original target function is replaced with a point mass at the mode (Müller (1999)). For 'large' J, the utility surface will become more peaked and simulations will cluster more tightly around the mode. An annealing schedule is not required, i.e., the same value of J may be used for all simulations, but this is not efficient for high dimensional problems (see Amzal et al. (2006)). For high dimensional problems a 'cooling' schedule may be used where J increases to $+\infty$, which is similar to decreasing the temperature to 0 in simulated annealing. Müller (1999) stated that for sufficiently large J, the sample mean of the simulated designs provides a good approximation of the optimal design, so that $h(\mathbf{d})$ does not have to be reconstructed. However, increasing J obviously increases the number of required computations.

Whilst the algorithm presented by Müller (1999) has 'theoretically appealing' properties, it has been found to have slow convergence in practice, particularly for situations where there are a large number of design variables for which this algorithm becomes inefficient (Stroud et al. (2001); Amzal et al. (2006)). Use of this algorithm has therefore mostly been restricted to up to four design variables (e.g., Bielza et al. (1999); Müller (1999); Stroud et al. (2001); Cook et al. (2008)). However, Sanso and Müller (1997) used a Bayesian design approach to look at the design problem of reducing a current system of 80 rainfall monitoring stations to around 40 stations, in a way that has the least reduction in the ability to make inferences about local rainfall and minimise cost. A simulation-based approach was used to solve the optimal design problem (Clyde et al. (1996)), but the design problem was reduced to a variable selection problem, which reduced the computational burden for the large amounts of data.

Several authors (e.g., Chaloner and Larntz (1989); Dror and Steinberg (2008); Woods et al. (2006); Gotwalt et al. (2009)) have found pseudo-Bayesian D-optimal designs for a large number of design points (up to 100 in some cases) using numerical search algorithms. However, these designs cannot be considered truly Bayesian, as they were simply robust to uncertainty in the model parameters $\boldsymbol{\theta}$ and their utility functions did not involve data. Bayesian utility functions, which involve data, would be much more computationally intensive for a large number of design points.

Stroud et al. (2001) used the simulation-based approach of Müller (1999) to determine the optimal sampling times for the anticancer agent, paclitaxel, using design criteria related to the total area of the curve, the time above a critical threshold value, and the sampling cost. In their study they considered up to three sampling times. Stroud et al. (2001) found that beyond 3 sampling times, their methods could not be performed within reasonable computation times. In their concluding statements, Stroud et al. (2001) suggested that for more than three sampling times, one could use a lower dimensional parameterisation of fixed multiple sampling times. To our knowledge, this method of using lower dimensional parameterisations has not been previously implemented.

A better and more rapid exploration of the design space is required. Also, for experimental design problems where the design space is a simplex (e.g., ordered sampling times), one should search for the multivariate or joint mode over the simplex rather than the marginal modes of the design variables, since the marginal modes may give misleading designs. However, for design problems that involve a large number of design points, the problem of finding the multivariate mode is more difficult than finding marginal modes.

In this paper we implement methods which enable one to use the above-mentioned simulation-based approaches for finding (sub) optimal designs for Bayesian design problems that require a large number of design points (sampling times) to be found, whose data are described by nonlinear models. We adopt the (previously-unused) approach presented by Stroud et al. (2001), which involves the use of lower dimensional parameterisations that consist of a few design variables which generate multiple design points. This avoids the need to search for a large number of optimal design points since one simply has to search over a few design variables, thus providing substantial computational savings. Also, it is much easier to obtain the multivariate mode for a few design variables than it is for a large number of design variables, and so use of these lower dimensional parameterisations enables the mode to be found with greater ease for design problems with a large number of design points. It should be noted that the designs which are generated by the lower dimensional parameterisations presented in this paper are sub-optimal, rather than optimal, which is a compromise of the computational savings achieved through these methods.

In Section 3.2 we describe how these simulation-based approaches are adapted for design problems that require a large number of design points to be found. Section 3.3 introduces the utility functions and design methodologies used in this work. In Section 3.4 we apply the algorithm to four examples. The paper concludes with a discussion in Section 3.5.

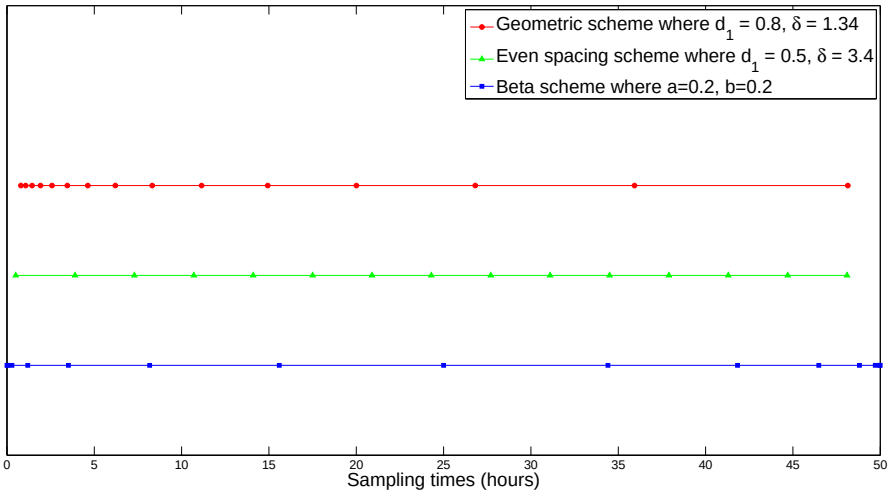


Figure 3.1: An illustration of some possible designs that one could obtain under the three lower dimensional parameterisation schemes investigated in this study.

3.2 Lower Dimensional Parameterisations

Here we extend the work of Clyde et al. (1996), Bielza et al. (1999) and Müller (1999) to enable a large number of design points to be found using Metropolis-Hastings MCMC simulations and Bayesian utility functions. Our approach is useful for design variables such as sampling times, that require multiple measures (many design points) to be taken at specific points that are separated from one another in the design space. Using an approach suggested by Stroud et al. (2001) (but previously not used), we investigate various lower dimensional parameterisations of multiple design points. For illustrative purposes we consider functions that consist of the first design point d_1, a spacing parameter δ, and an index s where $s = 1, ..., n_d$ and n_d is the number of design points. The first design point d_1 and the spacing parameter δ are proposed from appropriate distributions, say, a normal random walk. The remaining design points are then generated by using these proposed values of (d_1, δ) in conjunction with a lower dimensional parameterisation (see below). We also investigated a proposal scheme in which the design points were generated from the (evenly-spaced) percentiles of a beta distribution which was defined by two positive shape parameters (a, b). In this article, we have investigated the use of the following lower dimensional parameterisation schemes to generate the design points:

(a) $d_s = d_1\delta^{(s-1)}$, where $d_1 \geq 0, \delta > 1$ ('geometric scheme');
(b) $d_s = d_1 + \delta \times (s-1)$, where $d_1 \geq 0, \delta > 0$ ('even spacing scheme'); and
(c) Percentiles of a Beta(a, b) distribution, scaled to [0,T] - the design space, where $a, b > 0$ ('beta scheme').

An example of these schemes is provided in Figure 3.1.

Under these lower dimensional parameterisations of the design points, the Müller (1999) algorithm searched over two design variables (d_1, δ), or (a, b), depending on the scheme that was used. This avoids the need to search for a large number of optimal design

points and provides substantial computational savings. These lower dimensional parameterisation schemes were chosen with particular design problems/applications in mind (see Section 3.4), but alternative parameterisations may easily be incorporated into the algorithm, depending on the user's application.

3.3 Design Methodology

To solve the optimal design problems (equation (3.1)), we implemented the Metropolis-Hastings MCMC algorithm presented by Müller (1999), which is described in Algorithm 3.1, to perform simulation from $h(\cdot)$. For both examples present in this paper, the value of $J = 10$ was used. We found that this value of J provided us with sufficiently peaked utility surfaces. In Line 4 of Algorithm 3.1, a proposal distribution $q(\cdot)$ may be selected by the user. For the examples in this paper, a normal random walk was used for $q(\cdot)$. To adapt Algorithm 3.1 for design problems that require a large number of design points to be found, one simply uses one of the lower dimensional parameterisation schemes introduced in Section 3.2 in Line 4 of Algorithm 3.1. Use of these lower dimensional parameterisations are not restricted to the Müller (1999) MCMC algorithm, and one could use these schemes with other algorithms such as the algorithm proposed by Amzal et al. (2006).

Algorithm 3.1: MCMC algorithm for Bayesian optimal design

1 Initialise - set an initial design $\mathbf{d}^{(1)}$, simulate $(\boldsymbol{\theta}_j, \mathbf{y}_j)$ from $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}^{(1)}) = p(\boldsymbol{\theta})p(\mathbf{y}|\mathbf{d}^{(1)}, \boldsymbol{\theta})$ for $j = 1, ..., J$

2 Compute $U^{(1)} = \prod_{j=1}^{J} U(\mathbf{d}^{(1)}, \boldsymbol{\theta}_j, \mathbf{y}_j)$

3 **for** *i = 1 to iters* **do**

4 Generate a candidate design $\tilde{\mathbf{d}}$ from a proposal distribution $q(\cdot|\mathbf{d}^i)$ or a lower dimensional parameterisation scheme

5 Generate proposals for the parameters and simulate data $(\tilde{\boldsymbol{\theta}}_j, \tilde{\mathbf{y}}_j)$ from $p(\boldsymbol{\theta}, \mathbf{y}|\tilde{\mathbf{d}}) = p(\boldsymbol{\theta})p(\mathbf{y}|\tilde{\mathbf{d}}, \boldsymbol{\theta})$ for $j = 1, ..., J$

6 Compute $\tilde{U} = \prod_{j=1}^{J} U(\tilde{\mathbf{d}}, \tilde{\boldsymbol{\theta}}_j, \tilde{\mathbf{y}}_j)$

7 Calculate the MH acceptance probability, $a = \min(1, A)$ where

$$A = \frac{\tilde{U} \times q(\mathbf{d}^{(i)}|\tilde{\mathbf{d}})}{U^{(i)} \times q(\tilde{\mathbf{d}}|\mathbf{d}^{(i)})}$$

when the proposals for $(\boldsymbol{\theta}, \mathbf{y})$ come from the prior predictive distribution as above. Here $U^{(i)}$ and $\mathbf{d}^{(i)}$ are the current utility and design point values, respectively, and $\tilde{U}$ and $\tilde{\mathbf{d}}$ are the proposed utility and design point values, respectively.

8 Set $(\mathbf{d}^{(i+1)}, U^{(i+1)}) = (\tilde{\mathbf{d}}, \tilde{U})$ with probability a, and

9 $(\mathbf{d}^{(i+1)}, U^{(i+1)}) = (\mathbf{d}^{(i)}, U^{(i)})$ with probability $1 - a$.

10 **end for**

In this article we investigate Bayesian design criteria for efficient parameter estimation. More specifically, the design criteria that we investigate are: the Kullback-Leibler divergence between the prior and posterior, the inverse of the determinant of the posterior variance-covariance matrix, and a response variance criterion presented by Solonen et al.

(2012) (described below). Use of the Bayesian design criteria assumes that a Bayesian analysis will be performed on any data that is generated from the experimental design and involves integration over the parameter space, meaning that these design criteria are not a function of $\boldsymbol{\theta}$. For all of the utility functions mentioned below, we are interested in finding the optimal design $\mathbf{d}^*$, that maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$, with respect to the unknown data $\mathbf{y}$ and model parameters $\boldsymbol{\theta}$. Since the Bayesian utility functions $U(\mathbf{d}, \mathbf{y})$ do not typically have a closed form, we will use Monte Carlo methods (described below) to obtain suitable estimates of these utility functions. Here it is assumed that all parameters are equally important. If one were interested in precisely estimating only a subset of these parameters, then one could modify the existing utility functions, e.g., the Kullback-Leibler divergence between the parameters of interest, or one could use alternative utility functions.

The Kullback-Leibler divergence (KLD) (Kullback and Leibler (1951)) between the prior and posterior distributions is used as a design criterion when one is interested in maximising the expected information gain on the model parameters $\boldsymbol{\theta}$ due to performing an experiment at design points $\mathbf{d}$. The KLD between the posterior and prior has commonly been used in Bayesian experimental design (e.g., Bernardo and Smith (1994); Cook et al. (2008); Drovandi et al. (2013); Huan and Marzouk (2012, 2013)) and its expression is given by:

$$\begin{aligned} U(\mathbf{d}, \mathbf{y}) &= E_{\boldsymbol{\theta}|\mathbf{d},\mathbf{y}}(\log p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) - \log p(\boldsymbol{\theta})) \\ &= \int_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}) \log p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}) d\boldsymbol{\theta} - \log p(\mathbf{y}|\mathbf{d}) \end{aligned}$$

where $\mathbf{d}$ are the design points and $\mathbf{y}$ is the data yielded by the experiment (see Friel and Pettitt (2008)). To calculate the KLD, we used the approach given by Cook et al. (2008), in which the prior is discretised by drawing N_p values of $\boldsymbol{\theta}$ from it. In our applications we used $N_p = 10000$ or $N_p = 20000$, depending on the application. Importance sampling is then used to approximate the posterior distribution, where the prior is the importance distribution and the likelihood is used to calculate the importance weights. The following is used to estimate the KLD:

$$\widehat{U(\mathbf{d}, \mathbf{y})} = \sum_{i=1}^{N_p} \frac{p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_i)}{\sum_{l=1}^{N_p} p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_l)} \log p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_i) - \log \frac{1}{N_p} \sum_{i=1}^{N_p} p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_i)$$

where $p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_i) / \sum_{l=1}^{N_p} p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta}_l)$ are the importance weights. The discretisation of the prior was performed to facilitate the computation of the KLD. However, this estimator would perform poorly when uninformative priors are involved.

If one is interested in maximising the (joint) posterior precision (or minimising the posterior covariance) of all of the model parameters $\boldsymbol{\theta}$, then one could use the inverse of the determinant of the posterior variance-covariance matrix of the model parameters as the design criterion. The inverse of the determinant of the posterior variance-covariance

matrix of the model parameters is also known as the 'Bayesian D-posterior precision' (Drovandi et al. (2013)) and is given by:

$$U(\mathbf{d}, \mathbf{y}) = \frac{1}{\det(\text{cov}(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}))}.$$

This utility is estimated by finding the reciprocal of the determinant of the variance-covariance matrix of the $\boldsymbol{\theta}$ sample from the posterior. Importance sampling was again used to approximate the posterior distribution to calculate this utility function. If one were interested in maximising the precision of the marginal posterior distributions of the model parameters, then one should use the trace in the above equation instead of the determinant to obtain the Bayesian A-posterior precision. If the posterior distribution is multimodal, then use of the determinant of the posterior variance-covariance matrix utility function may be inappropriate and one should instead use the KLD utility function.

We will also investigate the utility presented by Solonen et al. (2012) where the next design point is placed where the prior variance of the mean response is largest. Therefore, one should use this utility function if interest lies in minimising the prior variance of the mean response. The utility is calculated by bringing in the observations one-at-a-time. The utility function is given by:

$$U(\mathbf{d}, \mathbf{y}) = \prod_{k=1}^{K} (\sigma^2 + \text{Var}_{\boldsymbol{\theta}|\mathbf{y}_{1:(k-1)}}(m_k(\boldsymbol{\theta}))), \tag{3.3}$$

where $m_k(\boldsymbol{\theta}) = E(y_k|d_k, \boldsymbol{\theta})$ and K is the number of observations. The expression $\text{Var}_{\boldsymbol{\theta}|\mathbf{y}_{1:(k-1)}}(m_k(\boldsymbol{\theta}))$ gives the variance of the mean response at d_k, given measurements $\mathbf{y}_{1:(k-1)}$ at points $\mathbf{d}_{1:(k-1)}$. The utility at d_k is evaluated using a weighted variance, where each simulated response is weighted based on the likelihood of previous simulated measurements, $p(\mathbf{y}_{1:(k-1)}|\mathbf{d}_{1:(k-1)}, \boldsymbol{\theta})$:

$$\widehat{\text{Var}}_{\boldsymbol{\theta}|\mathbf{y}_{1:(k-1)}}(m_k(\boldsymbol{\theta})) = \sum_{l=1}^{N_p} w_l (m_k(\boldsymbol{\theta}_l) - \overline{m_k}(\boldsymbol{\theta}))^2,$$

where

$$\overline{m_k}(\boldsymbol{\theta}) = \sum_{l=1}^{N_p} w_l m_k(\boldsymbol{\theta}_l)$$

and weights

$$w_l \propto p(\mathbf{y}_{1:(k-1)}|\mathbf{d}_{1:(k-1)}, \boldsymbol{\theta}_l) = \prod_{m=1}^{k-1} p(y_m|d_m, \boldsymbol{\theta}_l).$$

The weights are normalised to sum to one.

3.4 Examples

Here we apply our proposed design approach to four examples. The first example is a toy example which involves finding the optimal design points for a single regressor in a linear model. The second example involves selecting sampling times for a memory retention

exercise. The third example involves selecting the optimal blood sampling times for a pharmacokinetic (PK) model. The fourth example is a heat transfer problem that was also investigated by Solonen et al. (2012).

In the first two examples, we investigate the efficiency of the lower dimensional designs compared to unrestricted optimal designs (i.e., designs which were found by searching over n design variables and did not involve lower dimensional parameterisations). For the third and fourth examples, we are unable to investigate the efficiency of the designs due to the computational burden of finding the unrestricted designs.

The examples in this paper contain continuous design variables, but the approaches presented in this paper could be generalised for applications that require discrete design variables. For design problems that involve discrete design variables, one would have to choose a lower dimensional parameterisation that generates discrete values for the proposed values of the design points (Line 4 of Algorithm 3.1).

All simulations were performed on a desktop PC with a single 3.33-GHz Intel Core i5 processor.

3.4.1 Example 1: Linear regression

To investigate the efficiency of the lower dimensional designs, as a toy example, we considered a linear regression model with one design variable and n sampling points, $E[\mathbf{y}] = \beta_0 \cdot \mathbf{1} + \beta_1 \mathbf{x}$, where $\mathbf{y} \,|\, \boldsymbol{\beta}, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$. Here σ^2 is known and $\mathbf{I}$ is the $n \times n$ identity matrix. It was assumed that $\boldsymbol{\beta}|\sigma^2 \sim N(\bar{\boldsymbol{\beta}}, \sigma^2 \mathbf{R}^{-1})$, where $\mathbf{R}$ is a known 2×2 precision matrix.

The KLD was used as the utility function, and for this simple linear model, a closed form expression of the expected utility function is available: $U(\mathbf{d}) = \det(\mathbf{X}^T\mathbf{X} + \mathbf{R})$ (see Chaloner and Verdinelli (1995)). For this example, the precision matrix was assumed to be a 2×2 identity matrix.

The design space was restricted to $[0, 1]$ and numerical search algorithms (quasi-Newton) were used to find the optimal designs. The 'unrestricted optimal design' was found in which the numerical search algorithm searched over n design variables to find the optimal designs, i.e., no lower dimensional parameterisations were used. The beta scheme was then used to generate an n sampling point design, and the numerical algorithm searched for the optimal beta parameters (a, b).

We considered a design problem consisting of 11 design points ($n = 11$). The optimal unrestricted design consisted of 2 support points, $(0, 1)$, with 5 replicates on 0 and 6 replicates on 1. The optimal values of the shape parameters for the beta scheme were $(a^*, b^*) = (0.0038, 0.0032)$, which gives a "bath tub curve" for the beta density and also gave a two support point design with 5 replicates on 0 and 6 replicates on 1. Therefore, for this simple example, the lower dimensional scheme was able to generate fully efficient designs. However, for a more complex model, such as a nonlinear model, this may not hold true.

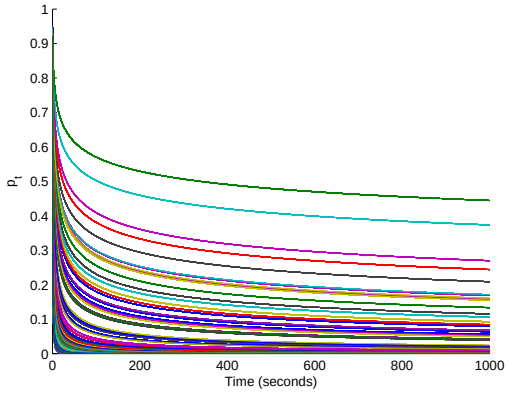


Figure 3.2: Mean response distributions for 100 simulated values of p_t for the memory retention example.

3.4.2 Example 2: Memory Retention

For this application we use one of the models for memory retention presented by Cavagnaro et al. (2010), but have changed the design problem slightly to that of a static design. In this hypothetical experiment, n trials are conducted, each of which assess memory retention at a single time point, and the data are aggregated across the trials. In each trial, the study subject is given a set of words to memorise, and after a particular lag time, the participant is asked to recall the list of words. A binary outcome is observed of whether or not the subject was able to correctly recall the initial list of words.

The model is given by:

$$p_t = \theta_0(t+1)^{-\theta_1}, \tag{3.4}$$

where p_t is the probability that the subject recalls the list of words at lag time t. The lag times are the design variables and the experimental aim is to determine the optimal set of lag times (across the experiments) for the purpose of parameter estimation. We are interested in finding a static design, rather than an adaptive design. For simplicity, only one lag time per trial was considered.

A fixed effects model is assumed. Uniform priors are assumed for the model parameters, $\theta_0 \sim \text{Unif}(0.9, 0.95)$ and $\theta_1 \sim \text{Unif}(0.1, 2)$. These priors were chosen based on the MLEs present in Rubin et al. (1999), and to enable the numerical integration (to calculate the expected utility) to be performed with ease. The prior predictive curves are displayed in Figure 3.2.

The KLD was used as the utility function. A very accurate approximation for the KLD was obtained by using numerical quadrature to integrate over the parameter space to obtain $U(\mathbf{d}, \mathbf{y})$. Since the response is binary, we were able to calculate the expected utility $U(\mathbf{d})$ exactly since it is possible to sum over all 2^n possible responses (for a reasonably small n).

The design space was restricted to $[0, 1000]$ seconds, where the lag times must be separated by a period of at least 10 seconds to enforce a separated design for illustrative purposes.

Proposal scheme	Optimal design	Utility function value
Unrestricted	$d^* = (3.67, 13.67, 23.67, 33.67, 43.67, 53.67, 63.67)$	0.43
Beta scheme	$d^* = (10.23, 20.23, 31.36, 44.40, 60.59, 82.54, 118.28)$, $(a^*, b^*) = (1.15, 18.38)$	0.42
Even spacing scheme	$d^* = (3.67, 13.67, 23.67, 33.67, 43.67, 53.67, 63.67)$ $(d_1^*, \delta^*) = (3.67, 10)$	0.43

Table 3.1: Optimal designs for the memory retention example.

Here we investigated 7 design points ($n = 7$). Numerical search algorithms (Nelder-Mead) were used to find the optimal designs. Unrestricted optimal designs were found, as well as designs generated by the beta and even spacing schemes. The geometric scheme was not used for this example as it was unable to generate designs that conformed to the design space restrictions.

For this example, the even spacing scheme was able to generate a fully efficient design, and the beta scheme was able to generate a design that was 98% efficient. Whilst we have been able to prove that these lower dimensional parameterisations are highly efficient for some design problems, the lower dimensional parameterisations may not be quite so efficient for other design problems and their efficiency is highly problem-dependent. Due to the computational burden involved, we are unable to determine the optimal unrestricted (15 design point) designs for Examples 3 and 4 (Sections 3.4.3 and 3.4.4 respectively). Instead, we take the high dimensional design space and project it onto a lower dimensional design space, and hope that the design generated by the optimal lower dimensional parameters is close to the optimal design.

3.4.3 Example 3: Sampling times for a PK study

Here we demonstrate our approach for the design problem of determining the optimal, or near optimal, placement of blood sampling times for a simple PK model. PK studies involve the administration of a specified quantity of a drug to an individual subject or group of subjects, and investigate the absorption, distribution and elimination of the drug and its metabolites (i.e., what the subjects' body does to the treatment). The kinetics of a drug cannot be directly observed in the study subjects. Instead, samples are taken from biological fluids such as blood, plasma, saliva or urine at specific times and the amount or concentration of drug and metabolites present in the sample are measured.

Compartmental models are commonly used to model the mean PK response of a subject, or group of subjects, to the drug and are derived by solving a series of ordinary differential equations that describe the time course of the drug's disposition (see Gibaldi and Perrier (1982)). Systematic and natural variation are present in the sample drug concentration data, and so statistical models are constructed by incorporating error terms into the compartmental model to account for this. These error terms are often heteroscedastic and various approaches for modelling the error structures have been suggested in the literature (e.g., Lunn et al. (2002); Davidian and Giltinan (2003)), most of which involve specifying the error in terms of some variance parameter, say σ^2, and a variance function $g(\boldsymbol{\theta}, \boldsymbol{\lambda}; \mathbf{t})$

which depends on the model parameters $\boldsymbol{\theta}$, dispersion parameters $\boldsymbol{\lambda}$ and sampling times $\mathbf{t}$. However, it is often difficult to accurately estimate PK parameters on an individual level, since it is impossible to take a large number of samples from each of the study subjects, and consequently only sparse individual samples are available. Thus, the planning of the timing and number of samples is of critical importance, so as to gain accurate estimates of the parameters but also prevent physical and mental strain on the study subjects.

For our motivating example, we chose to determine the (near) optimal blood sampling times for a one-compartment, first-order absorption and elimination, fixed effects PK model. This model does not account for individual variability. The model consists of three parameters: the volume of distribution V, which is a theoretical volume that a drug would have to occupy to provide the same concentration as is currently present in the blood plasma (if the drug were uniformly distributed); the first-order absorption rate constant k_a; and the first-order elimination rate constant k_e. If y_t denotes the observed concentration at time t following the administration of the drug, then the model may be given by:

$$y_t = \frac{D}{V}\frac{k_a}{k_a - k_e}(\exp(-k_e t) - \exp(-k_a t)) \cdot (1 + \epsilon_{1t}) + \epsilon_{2t}. \tag{3.5}$$

It is assumed that a single fixed dose $D = 400$ units is administered at the beginning of the experiment. It was also assumed that

$$\log \boldsymbol{\theta} \sim N\left[\begin{pmatrix} \log(1) \\ \log(0.1) \\ \log(20) \end{pmatrix}, \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.05 & 0 \\ 0 & 0 & 0.05 \end{pmatrix}\right],$$

where $\boldsymbol{\theta} = (k_a, k_e, V)$, $k_a > k_e$, $\epsilon_{1t} \sim N(0, \sigma^2_{\text{prop}})$, $\sigma^2_{\text{prop}} = 0.01$, $\epsilon_{2t} \sim N(0, \sigma^2_{\text{add}})$, $\sigma^2_{\text{add}} = 0.1$ and $t \in [0, 24]$. The priors were obtained using similar values that were obtained from other PK design studies in the literature (e.g., Duffull et al. (2012); McGree et al. (2012a)) and from examining the mean response (Figure 3.3).

Here our design points $\mathbf{d}$ are the sampling times $\mathbf{t}$. It is not practically feasible to take more than one blood sample at a time (replication) in PK studies, and so restrictions were placed on the lower dimensional parameterisations whereby the sampling times generated must be at least 10-15 minutes apart.

This model assumes that the components of $\boldsymbol{\theta}$ are independent with equal uncertainty of prior specification. The values of the proportional and additive variance terms (variance of ϵ_{1t} and ϵ_{2t} respectively) were chosen to give plausible models.

For the PK example, we investigated all three lower dimensional parameterisation schemes mentioned in Section 3.2 (geometric, even spacing and beta schemes) and used these schemes to generate 15 sampling times that occurred between 0 and 24 hours. The Müller (1999) algorithm was used to search over two design variables, (d_1, δ), or (a, b), depending on the scheme which was used.

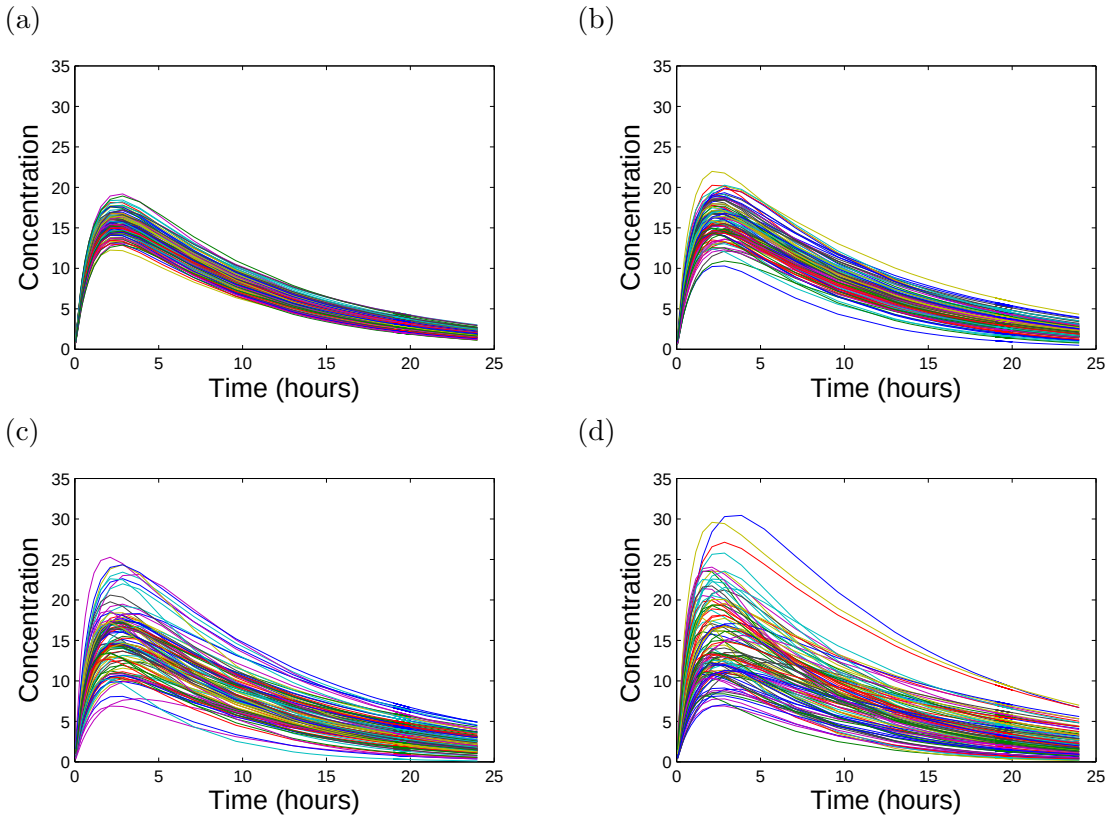


Figure 3.3: Mean response distributions for 100 simulated values of y (the concentration) where the variance-covariance matrix for $\boldsymbol{\theta}$ is varied using the following values: (a) $0.01 \cdot \mathbf{I}$, (b) $0.02 \cdot \mathbf{I}$, (c) $0.05 \cdot \mathbf{I}$, and (d) $0.1 \cdot \mathbf{I}$, where $\mathbf{I}$ is a 3×3 identity matrix.

The utility function developed by Solonen et al. (2012) assumes a constant variance (equation (3.3)). However, our PK model has a non-constant variance since the model contains both additive and proportional error (which depends on the mean at time t). To account for this heterogeneity, we generalised the utility function presented by Solonen et al. (2012) by using the unconditional variance of the response rather than the variance of the mean of the response. Our generalised utility function is given by:

$$\begin{aligned} U(\mathbf{d}, \mathbf{y}) &= \prod_{k=1}^{K} \text{Var}(\mathbf{y}_k | \mathbf{y}_{1:(k-1)}) \\ &= \prod_{k=1}^{K} E_{\mu_k | y_{1:(k-1)}}(\text{Var}(y_k | \mu_k) + \text{Var}(\mu_k)) \\ &= \prod_{k=1}^{K} (\sigma^2_{\text{add}} + \sigma^2_{\text{prop}} \cdot E_{\mu_k | y_{1:(k-1)}}(\mu_k^2) + \text{Var}_{\mu_k | y_{1:(k-1)}}(\mu_k)), \end{aligned} \tag{3.6}$$

where $E(\mathbf{y}|\mu) = \mu$. The idea behind this generalised version of Solonen et al.'s (2012) utility function is to place the next design point where the prior predictive variance of $\mathbf{y}$ is largest. We will term this utility function the 'prior predictive response variance' utility. In addition to this utility function, we also used the KLD between the prior and posterior distributions, and the inverse of the determinant of the posterior variance-covariance matrix.

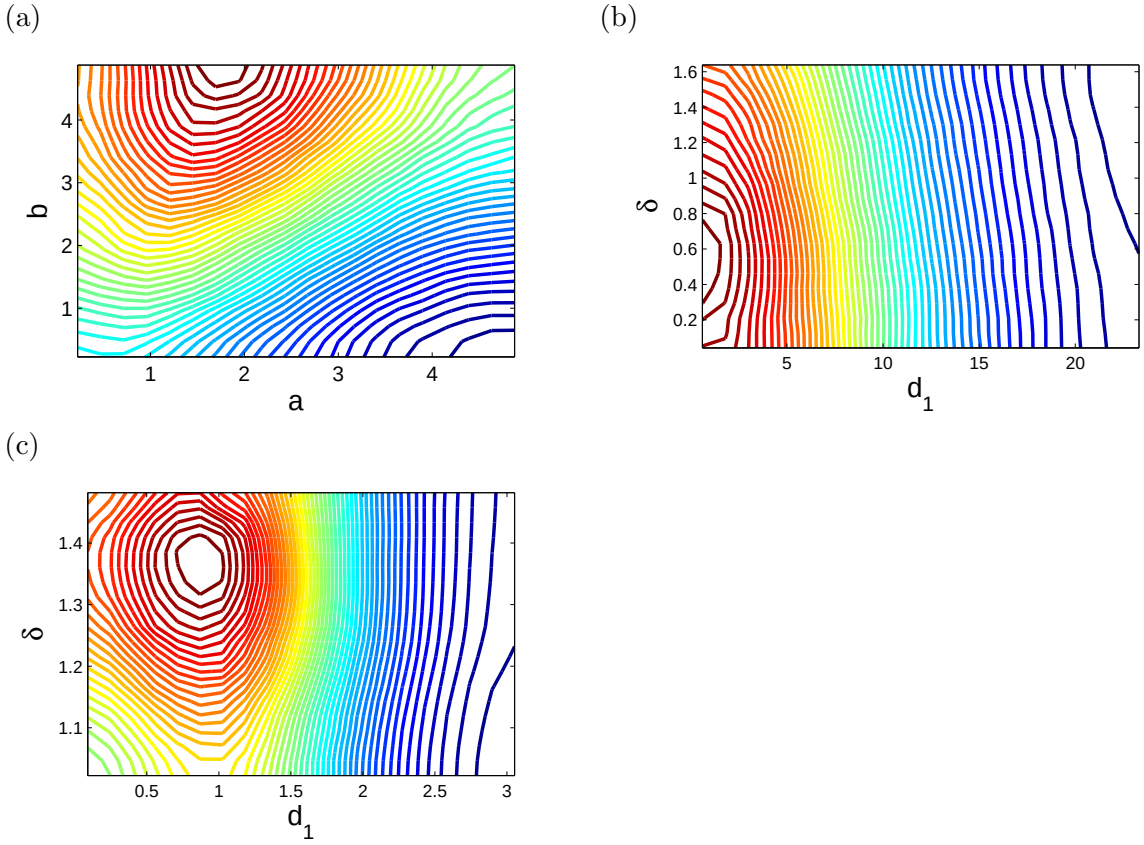


Figure 3.4: Two-dimensional contour plots of the expected utility surface (prior predictive response variance) for the various lower dimensional parameterisations for the PK example (a) Beta scheme, (b) Even spacing scheme, and (c) Geometric scheme.

The optimal designs were found by searching for the bivariate mode of the multivariate normal kernel smoothing density estimates of the design variables (see Cook et al. (2008); Drovandi and Pettitt (2013)). The densities were estimated based on (thinned) samples of the design variables obtained during the MCMC. Two-dimensional contour plots of the thinned posterior samples of the design variables for the various lower dimensional parameterisations that were generated using the prior predictive response variance as the utility function are displayed in Figure 3.4. The marginal densities of the design variables were also examined, but these were found to be misleading in some situations since the joint densities were found to have a 'banana' shape. The utility function values for the optimal designs (Tables 3.2, 3.3) were calculated using Monte Carlo integration (equation (3.2)) with $M = 10000$.

For each of the different lower dimensional proposal schemes, it was found that the designs were somewhat different (but not drastically different) across the different utilities, (Figure 3.5; online figures are in colour). For the geometric scheme, the optimal designs were approximately equivalent across the different utilities. Within each of the different utility functions, the spread of the designs was somewhat similar, apart from the prior predictive response variance design criterion. The beta scheme produced designs which gave the highest utility function values (out of the three lower dimensional parameterisation schemes) for the KLD utility and the prior predictive response variance utility. The

Design criterion	Proposal scheme	Utility function value (95% CI)**	Optimal design	Run time (hr)
KLD	Beta	**3.13 (3.11, 3.15)**	$(a^*, b^*) = (0.7, 1.2)$	2
	Even spacing	3.06 (3.04, 3.08)	$(d_1^*, \delta^*) = (0.3, 1.6)$	
	Geometric	3.01 (3.00, 3.03)	$(d_1^*, \delta^*) = (0.94, 1.25)$	
1/det(posterior var-cov)	Beta	$13.5(4.42, 22.6) \times 10^6$	$(a^*, b^*) = (0.8, 2.07)$	2
	Even spacing	$7.73(6.21, 9.25) \times 10^6$	$(d_1^*, \delta^*) = (0.73, 1.13)$	
	Geometric	$\mathbf{23.90(4.33, 43.47) \times 10^6}$	$(d_1^*, \delta^*) = (0.93, 1.25)$	
Prior predictive response variance	Beta	$\mathbf{25.5(25.2, 25.8) \times 10^{14}}$	$(a^*, b^*) = (1.58, 3.8)$	1
	Even spacing	$3.94(3.93, 3.95) \times 10^{14}$	$(d_1^*, \delta^*) = (2.2, 0.8)$	
	Geometric	$7.02(6.82, 7.22) \times 10^{14}$	$(d_1^*, \delta^*) = (0.94, 1.25)$	

** The proposal schemes which produced the designs that gave the highest utility values have been boldfaced for each utility function.

Table 3.2: Utility function values for the various proposal schemes for the PK example.

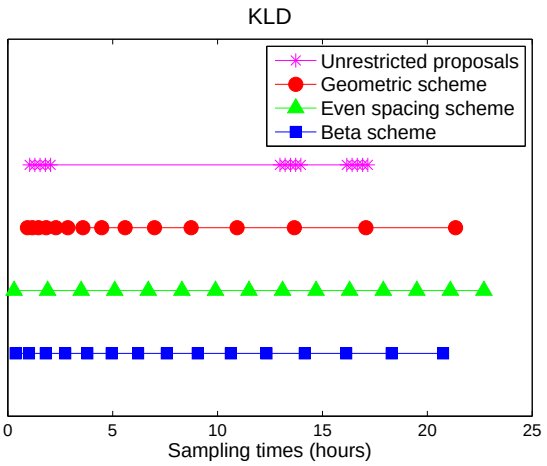
geometric scheme produced designs which gave the highest utility values for the inverse of the determinant of the posterior variance-covariance matrix utility utility (Table 3.2).

To assess the validity of our lower dimensional parameterisation schemes, we also searched for optimal designs for a three design (support) point problem using the Müller (1999) algorithm, where the design variables were (d_1, d_2, d_3), i.e., the three sampling times for the PK study. This involved searching for three optimal sampling times which were generated in the MH algorithm via normal random walks, and did not involve the use of the lower dimensional parameterisation schemes. To obtain the multivariate mode of (d_1, d_2, d_3), we used the approach described by Drovandi and Pettitt (2012) which involved the use of a multivariate Gaussian smoothing kernel (see, for example, Wand and Jones (1994)) on the samples of the design variables from the MCMC runs (see also Cook et al. (2008)).

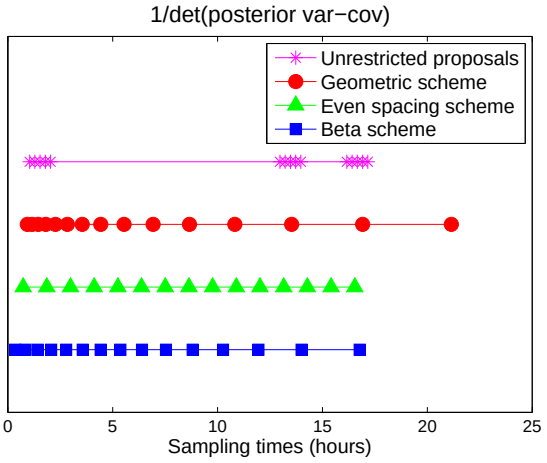
Once these three sampling times were determined, 5 replicates were placed on each of these support points. Since true replication is not practically feasible for a PK study, these 'replicates' were separated by a time interval of 15 minutes (sampling times 15 and 30 minutes before the support point, and sampling times 15 and 30 minutes after the support point). These designs were similar for the KLD and inverse of the determinant of the posterior variance-covariance matrix, whereas the designs for the prior predictive response variance differed somewhat from the other utility functions (Figure 3.5). The utilities for both the three design point and replicate (15 design point) problems were calculated (Table 3.3) and compared to utility values that were obtained using the three lower dimensional parameterisation schemes (Table 3.2).

For all utility functions investigated, it was found that the fifteen sampling times that were obtained via either the lower dimensional parameterisation schemes or by replication gave higher utility values than the optimal three sampling times that were obtained, as was expected. Also, for each of the utility functions, it was found that all of the lower dimensional parameterisation schemes gave designs that produced higher expected utility values than the designs that were obtained via replication. The simulations which

(a)



(b)



(c)

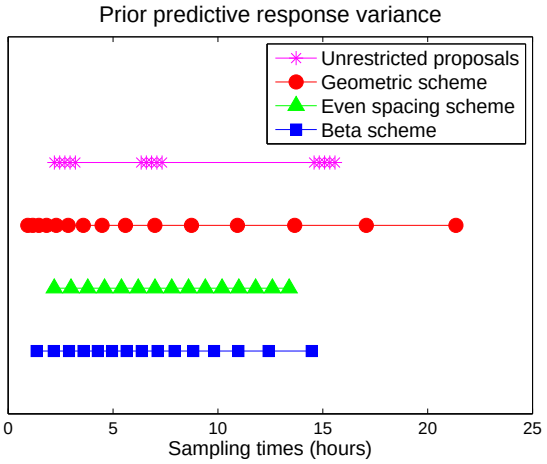


Figure 3.5: Sampling times for the PK example generated under the three parameterisation schemes using (a) the KLD, (b) the inverse of the determinant of the posterior variance-covariance matrix, and (c) the prior predictive response variance as the utility function. Displayed in the top row of each plot are the optimal replicate designs that were obtained by searching for three optimal design points, using unrestricted proposals (i.e., no lower dimensional parameterisations), and placing 5 replicates on each of these points.

Design criterion	No. design points	Utility function value (95% CI)	Optimal design	Run time (hr)
KLD	3	2.29 (2.28, 2.31)	$(d_1^*, d_2^*, d_3^*) = (1.54, 13.47, 16.66)$	5
1/det(posterior var-cov)	3	$2.17(2.13, 2.20) \times 10^5$	$(d_1^*, d_2^*, d_3^*) = (1.54, 13.47, 16.66)$	5
Prior predictive response variance	3	646.56 (640.25, 652.87)	$(d_1^*, d_2^*, d_3^*) = (2.71, 6.84, 15.08)$	2
KLD	15 (replicates**)	2.98 (2.96, 3.00)		
1/det(posterior var-cov)	15 (replicates**)	$5.26(4.93, 5.59) \times 10^6$		
Prior predictive response variance	15 (replicates**)	$3.32(3.31, 3.33) \times 10^{13}$		

** Once optimal designs were found for the unrestricted proposals for the 3 design point problem using the Müller (1999) algorithm, 5 replicates were placed on each of the support points and the utilities for the resulting 15 design point designs were calculated.

Table 3.3: Utility function values for a three design point problem, and for designs consisting of replicates of the three optimal design points (15 design points), under various utility functions for the PK example.

Utility function 1 $U_1(\mathbf{d}, \mathbf{y})$	Utility function 2 $U_2(\mathbf{d}, \mathbf{y})$	$U_2(\mathbf{d_{U_1}})/U_2(\mathbf{d}^*)$ (95% CI)
KLD	1/det(posterior var-cov)	1.01 (0.99, 1.03)
Prior predictive response variance	1/det(posterior var-cov)	0.36 (0.35, 0.40)
1/det(posterior var-cov)	KLD	0.95 (0.95, 0.95)
Prior predictive response variance	KLD	0.84 (0.83, 0.84)
KLD	Prior predictive response variance	0.00 (0.00, 0.00)
1/det(posterior var-cov)	Prior predictive response variance	0.11 (0.11, 0.11)

Table 3.4: Comparison of the optimal designs for each utility function ($U_1(\mathbf{d}, \mathbf{y})$) evaluated at the other utility functions ($U_2(\mathbf{d}, \mathbf{y})$) for the PK example.

made use of the lower dimensional parameterisation schemes were found to be much less computationally intensive than simulations which searched for three design points.

A quantitative comparison of the optimal designs and utility functions was carried out, in which the optimal design obtained from one utility function (e.g., KLD) was input into another utility function (e.g., 1/det(posterior var-cov)) and the resulting utility function value was compared to value of the utility function (e.g., 1/det(posterior var-cov)) evaluated at its optimal design. This is similar to the idea of 'D-efficiency' present in classical design. This quantitative comparison is given in Table 3.4 for the PK example:

Here 'utility function 1' is the utility function that the optimal design originally came from, and 'utility function 2' is the 'new' utility function that the design is being input into. $U_2(\mathbf{d_{U_1}})$ is the value of the 'new' utility function evaluated at the design that came from utility function 1. $U_2(\mathbf{d}^*)$ is the value of utility function 2 evaluated at its optimal design (from Table 3.3). The ratio of these quantities, $U_2(\mathbf{d_{U_1}})/U_2(\mathbf{d}^*)$, is evaluated in the third column of Table 3.4 to give a measure of efficiency of the design that was optimal for another utility function.

For the majority of the utility functions, the designs that were optimal for one utility function did not perform as well when input into another utility function (compared to that utility function's own optimal design). This is not surprising since the designs were generated under different objectives. In particular, the optimal designs for KLD and the

inverse of the determinant of the posterior variance-covariance matrix did not perform well when input into the prior predictive response variance utility function. When the optimal designs for the prior predictive response variance were input into the KLD and inverse of the determinant of the posterior variance-covariance matrix utilities, it was found that these designs did not perform as well as those utility functions' optimal designs. The KLD and inverse of the determinant of the posterior variance-covariance matrix utility functions were approximately equally efficient at designing for precisely estimating parameters (the KLD was slightly more efficient as indicated by the ratio values being just above 1). This may indicate that the prior predictive response variance utility function does not perform as well as the other utility functions for designing for efficient parameter estimation.

3.4.4 Example 4: Exothermic/Cooling study

For this application we use the same model as that which is presented by Solonen et al. (2012), but have changed the design problem slightly.

A glass of liquid that has an initial temperature of $T(0)$ is cooled or heated by water that is at a temperature of T_w. The surrounding air has a temperature of T_a. The model for the liquid temperature is given by

$$\frac{dT}{dt} = -\frac{k_w A_w}{Mc}(T - T_w) - \frac{k_a A_a}{Mc}(T - T_a), \tag{3.7}$$

where k_w and k_a are the heat transfer coefficients through glass and through the air-liquid interphase, respectively; A_w and A_a are the areas of water-liquid and air-liquid interfaces, respectively; c is the specific heat capacity; and M is the mass of liquid. We placed a restriction on the prior design space where $T(0) > T_w$ so that the mean response function decreased with time (i.e., was a cooling curve). The observations are given by the solution to equation (3.7), $T(t)$, which is given by equation (3.8), plus normally distributed, independent error with a constant variance $\sigma^2 = 0.33$:

$$T(t) = \frac{A_a T_a k_a - C_1 e^{-t(A_a k_a + A_w k_w)/Mc)} + A_w T_w k_w}{A_a k_a + A_w k_w}, \tag{3.8}$$

where $C_1 = -T(0)(A_a k_a + A_w k_w) + A_a T_a k_a + A_w T_w k_w$. The value of $\sigma^2 = 0.33$ is the same as that which was used by Solonen et al. (2012).

The purpose of the experiment is to estimate $\boldsymbol{\theta} = (k_w, k_a)$. In Solonen et al.'s (2012) work, the design problem was to determine the optimal $T(0)$ and T_w values. Measurements of the temperature were taken every two minutes, stopping at 20 minutes.

We extend the work of Solonen et al. (2012) by looking for 'closer to optimal' times at which to take the 10 temperature measurements, rather than simply taking the measurements at two minute intervals. Here the design parameters to be optimised are $\mathbf{d} = (T(0), T_w, d_1, \delta)$, or $\mathbf{d} = (T(0), T_w, a, b)$, depending on the lower dimensional parameterisation that is used. The design space for the sampling times was extended from $[0, 1200]$ seconds to $[0, 2000]$ seconds. The 10 sampling times were generated using the

lower dimensional parameterisation schemes mentioned in Section 3.2 (geometric, even spacing, and beta).

It was assumed that one set of measurements had already been taken at $(T(0), T_w) = (23, 5)$ and the parameter estimates that resulted from the MCMC simulations that fitted the model to the data were used to construct the prior distribution for $\boldsymbol{\theta}$.

For this example, we investigated all three Bayesian utility functions discussed in Section 3.2. We used 20000 particles to estimate our utility functions via importance sampling. The run time for these examples was approximately 0.5 hr when the mean response variance utility was used (for each of the proposal schemes) and 1.5 hrs for when the KLD and inverse of the determinant of the posterior variance-covariance matrix were used as the utility functions.

For all three design criteria, and all three proposal schemes, the optimal temperatures were found to occur at either extreme of the design space: $(T(0), T_w) = (60, 4)$ (i.e., hot liquid, cold water). To reduce the computational time, we set $(T(0), T_w) = (60, 4)$ and re-ran the MCMC simulations to find the optimal designs for two design variables (d_1, δ) or (a, b), depending on the scheme that was used. Two-dimensional contour plots of the thinned posterior samples of the design variables for the various lower dimensional parameterisations that were generated using the mean response variance as the utility function are displayed in Figure 3.6.

The sampling times generated by the different utility functions and proposal schemes are given in Figure 3.7 (online figures are in colour), and the utility function values for the different proposal schemes are given in Table 3.5. The utility function values for the optimal designs (Table 3.5) were calculated using Monte Carlo integration (equation (3.2)) with $M = 20000$.

For all three utility functions, the geometric scheme produced designs which gave the highest utility values, followed by the even spacing scheme, and the beta scheme (which gave the lowest utility values) (Table 3.5). The designs that resulted from each of the lower dimensional parameterisation schemes were found to be quite similar across the different utility functions, and produced higher utility function values than the original sampling times used by Solonen et al. (2012) (where samples were taken every 2 minutes from 0 to 20 minutes). Within each of the utility functions, the designs were somewhat similar across the different lower dimensional parameterisations in that they were fairly clustered and occurred around a similar region of the design space, with the beta scheme giving a wider coverage than the other two schemes. None of the parameterisation schemes generated designs that were spread over the entire design space.

A quantitative comparison of the optimal designs and utility functions was carried out in the same manner as for Section 3.4.3.

For all of the utility functions, the designs that were optimal for one utility function did not perform as well when input into another utility function (compared to that utility function's own optimal design). This is most noticeable when the optimal designs from

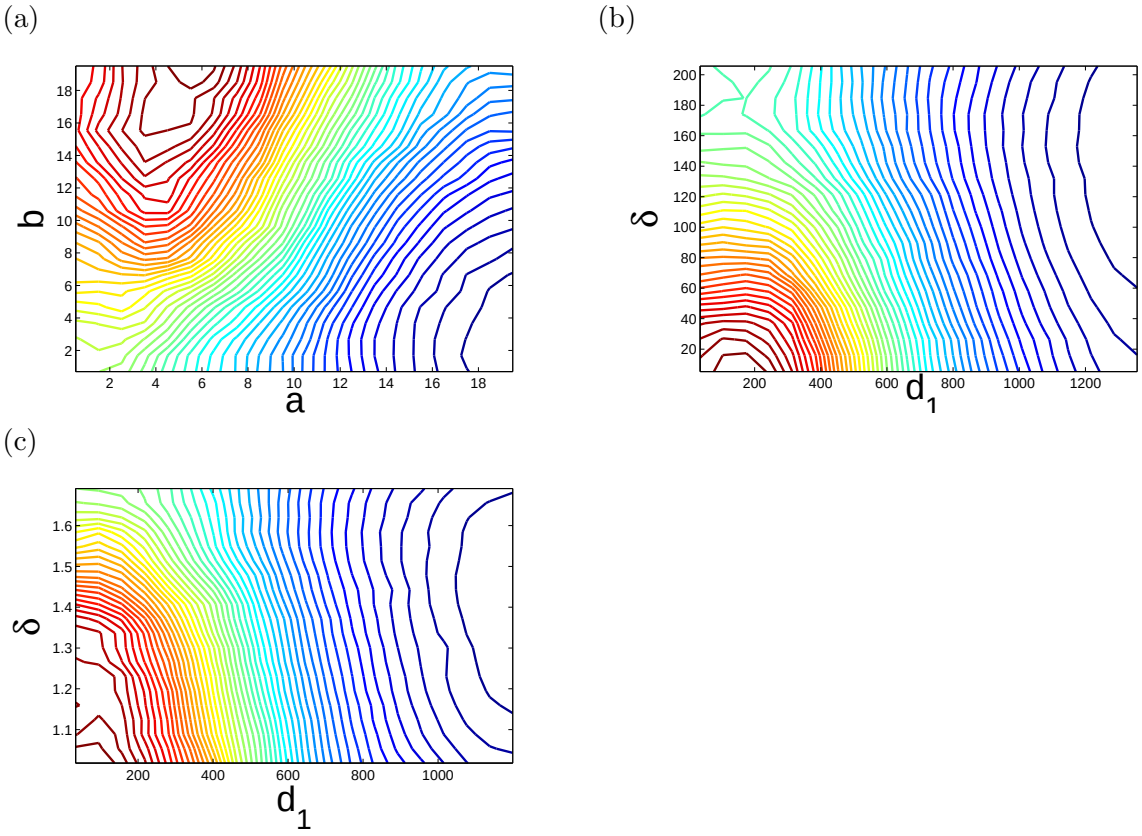


Figure 3.6: Two-dimensional contour plots of the expected utility surface (mean response variance) for the various lower dimensional parameterisations for the cooling example (a) Beta scheme, (b) Even spacing scheme, and (c) Geometric scheme.

Design criterion	Proposal scheme	Log utility function value (95% CI)**	Optimal design
KLD	Even spacing	0.50 (0.50, 0.51)	$(d_1^*, \delta^*) = (271.97, 34)$
	Beta	0.45 (0.45, 0.46)	$(a^*, b^*) = (4.73, 14.14)$
	Geometric	**0.54 (0.54, 0.55)**	$(d_1^*, \delta^*) = (205, 1.11)$
	Original design	0.35 (0.35, 0.36)	NA
1/det(posterior var-cov)	Even spacing	60.55 (60.55, 60.55)	$(d_1^*, \delta^*) = (267.75, 36.44)$
	Beta	60.46 (60.46, 60.47)	$(a^*, b^*) = (4.72, 14.83)$
	Geometric	**60.75 (60.75, 60.75)**	$(d_1^*, \delta^*) = (196.72, 1.12)$
	Original design	57.41 (57.40, 57.42)	NA
Mean response variance	Even spacing	3.45 (3.45, 3.45)	$(d_1^*, \delta^*) = (288.89, 31.29)$
	Beta	3.34 (3.34, 3.34)	$(a^*, b^*) = (4.59, 14.14)$
	Geometric	**3.73 (3.73, 3.73)**	$(d_1^*, \delta^*) = (219.58, 1.07)$
	Original design	3.34 (3.34, 3.34)	NA

** The proposal schemes which produced the designs that gave the highest utility values have been boldfaced for each utility function.

Table 3.5: Utility function values for the various proposal schemes and the original design used by Solonen et al. (2012) for the cooling example.

Utility function 1 $U_1(\mathbf{d}, \mathbf{y})$	Utility function 2 $U_2(\mathbf{d}, \mathbf{y})$	$U_2(\mathbf{d_{U_1}})/U_2(\mathbf{d}^*)$
KLD	1/det(posterior var-cov)	0.94 (0.94,0.94)
Mean response variance	1/det(posterior var-cov)	0.95 (0.95, 0.95)
1/det(posterior var-cov)	KLD	0.98 (0.98, 0.98)
Mean response variance	KLD	0.88 (0.88, 0.89)
KLD	Mean response variance	0.77 (0.77, 0.77)
1/det(posterior var-cov)	Mean response variance	0.76 (0.76, 0.76)

Table 3.6: Comparison of the optimal designs for each utility function ($U_1(\mathbf{d}, \mathbf{y})$) evaluated at the other utility functions ($U_2(\mathbf{d}, \mathbf{y})$) for the cooling example.

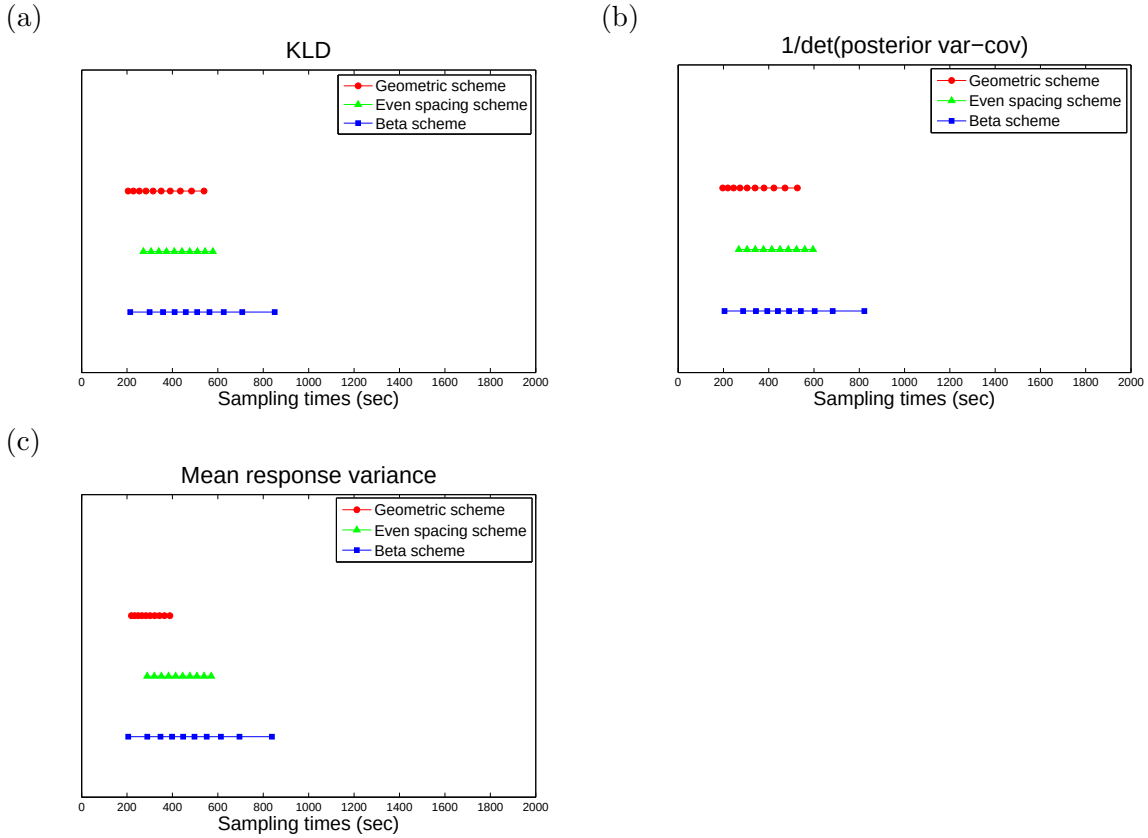


Figure 3.7: Sampling times (in seconds) for the cooling example generated under the three lower dimensional parameterisation schemes using (a) the KLD, (b) the (inverse of the) determinant of the posterior variance-covariance matrix, and (c) the mean response variance as the utility function.

the KLD and the inverse of the determinant of the posterior variance-covariance matrix utilities were input into the mean response variance utility. This suggests that the mean response variance utility was not as efficient at designing for precisely estimating parameters as the other utilities were.

3.5 Discussion

In this paper we have introduced a method for solving Bayesian design problems which require a large number of designs points to be found, through incorporation of lower dimensional parameterisations into existing stochastic optimisation algorithms. These lower dimensional parameterisations consisted of a few design variables, which were then input into various functions to generate multiple design points. This approach was found to have substantial computational savings since one simply has to search over a few design variables, rather than a large number of design variables. Also, it was much easier to obtain the multivariate mode for a few design variables than for a large number of design variables. However, it should be stressed that the designs presented in this paper are not optimal but *near* optimal, which is a compromise of the computational savings achieved through these methods.

Our approach is useful for design variables (e.g., sampling times) that require multiple measures to be taken at specific points that are separated from one another in the design space (although as shown in a toy example, the lower dimensional parameterisations can produce simple replicated designs). This approach does not overcome the problem of having a large number of different design variables (e.g., temperatures, pressures), and further research should be conducted for solving this design problem.

The functions that were used in this paper to generate the designs were purely illustrative for the examples chosen for this paper. There are many other functions and transformations available that one could choose to generate their design points and choice of an appropriate function would depend on the application of interest. To determine which function is most appropriate for a particular application, one could run several different parameterisations in parallel on different CPUs and choose the function which gives rise to the design with the highest utility.

We found that the beta proposal scheme, where the designs come from the (evenly-spaced) percentiles of a beta distribution, gave quite flexible designs, and so this function may be appropriate in many situations to generate a large number of design points. One could also extend the beta proposal scheme to propose from a generalised beta distribution (e.g., Sepanski and Kong (2007)), which may offer further flexibility in constructing the designs. One could also include another design variable in the parameterisation of the beta proposal scheme that determines the optimal percentiles of the beta distribution to use, e.g., percentile $= 100(\left(\frac{i}{n}\right)^{\alpha})$ where α is an additional design variable to search over.

In our PK example, the designs which were generated by the lower dimensional parameterisations gave higher utility values than the replicate designs. This is useful since it may not be feasible to take replicate designs for studies in which one is interested in determining the optimal sampling times. However, this may not always be the case and replicate designs may be preferred in some applications. If replicate designs are practically feasible for the experiment of interest, then one may wish to instead generate a large number of design points based on the estimated weights of a continuous or approximate design, which also falls into our framework of lower dimensional parameterisations. We will be investigating this in future work.

A fixed number of sampling times were assumed for the examples used in this study, so that we could demonstrate our methodology for generating many design points. The number of sampling times used in this study may not be optimal and future studies may wish to investigate the optimal number of design points for their applications. For experiments where cost is of little importance and the utility function does not contain a cost function, in general, the more design points there are the better.

Importance sampling did not perform particularly well (low effective sample sizes were often obtained) when it was used to estimate the Bayesian utility functions (especially for the KLD) for nonlinear models with a large number of observations. A large number of samples (10000-20000) was required to ensure that the utility was well estimated. To estimate a utility well, a certain number of samples is required. This number of samples

may vary across the utilities. KLD is harder to estimate in general since it involves the estimation of the evidence. The mean response variance utility may require less importance samples since the utility is estimated by bringing in observations one-at-a-time. For another paper we are investigating alternative methods for estimating Bayesian utility functions, such as Laplace approximations.

Acknowledgments

E.G. Ryan was supported by an APA(I) Scholarship which came from an ARC Linkage Grant with Genentech (LP0991602). The work of C.C. Drovandi and A.N. Pettitt was also supported by an ARC Discovery Project (DP110100159).

The authors would like to thank Dr James McGree from Queensland University of Technology for his helpful comments throughout the process of writing this paper. The authors would also like to thank the referees and the associate editor for their insightful comments and suggestions throughout the review process of this paper.

Statement of Authorship for Chapter 4

This chapter has been written as a journal article. The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
- They agree to the use of the publication in the students thesis and its publication on the QUT ePrints database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for the associated publication is: **Ryan, E.G.**, Drovandi, C.C., and Pettitt, A.N. (2014). Fully Bayesian Designs for Nonlinear Mixed Effects Models. Submitted to *Computational Statistics and Data Analysis*.

Contributor	Statement of contribution
Elizabeth Ryan	Wrote all Matlab code required for the manuscript, performed all computations in the manuscript, interpreted and reported the results, constructed all figures presented in the manuscript, wrote the manuscript, and acted as the corresponding author
Signature and Date:	
Christopher Drovandi	Assissted in the writing of the Matlab code, directed the research and proofread the manuscript.
Tony Pettitt	Directed the research and proofread the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all co-authors confirming their certifying authorship.

Name	Signature	Date

Chapter 4

Fully Bayesian Designs for Nonlinear Mixed Effects Models

ABSTRACT

In this paper, we present fully Bayesian experimental designs for nonlinear mixed effects models, in which we develop simulation-based optimal design methods to search over both continuous and discrete design spaces. Although Bayesian inference has commonly been performed on nonlinear mixed effects models, there is a lack of research into performing Bayesian optimal design for nonlinear mixed effects models that require searches to be performed over several design variables. This is likely due to the fact that it is much more computationally intensive to perform optimal experimental design for nonlinear mixed effects models than it is to perform inference in the Bayesian framework. In this paper, the design problem is to determine the optimal number of subjects and samples per subject, as well as the (near) optimal urine sampling times for a population pharmacokinetic study in horses, so that the population pharmacokinetic parameters can be precisely estimated, subject to cost constraints. The optimal sampling strategies, in terms of the number of subjects and the number of samples per subject, were found to be substantially different between the examples considered in this work, which highlights the fact that the designs are rather problem-dependent and can be addressed using the methods presented in this paper.

KEYWORDS: Bayesian optimal design; Nonlinear mixed effects models; Population design; Sampling strategies; Stochastic optimisation.

4.1 Introduction

4.1.1 Background

Nonlinear mixed effects models (NLMEMs) are commonly used to model data in which heterogeneity exists between study subjects. For example, population pharmacokinetic (PK) studies investigate the disposition of a drug within a large sample of subjects. For an extensive overview of NLMEMs, as well as general theoretical developments and examples, see Racine-Poon (1985) and Davidian and Giltinan (1995). NLMEMs require the mean value of the population parameters to be estimated, as well as their inter-individual variability. The experimental design, which is usually under the control of the investigator, is responsible for determining the quality of the analyses of the data modelled by the NLMEM. These models, which are also known as the *population approach*, can allow

for a sparse sampling design where only a few data points are available per individual, but a large number of individuals are included in the study. This is useful for studies in which the experimenters are interested in collecting informative data to obtain precise parameter estimates, but the number of samples per subject is limited due to ethical, physiological, time or cost constraints.

As the use of NLMEMs for modelling data from population studies has increased, so has the importance of optimally designing population studies so that accurate and precise estimates of the population parameters can be obtained (e.g., Mentré et al. (1997); Retout and Mentré (2003); Han and Chaloner (2004)). In this paper we are interested in (static) experimental designs that are optimal for the estimation of the population parameters in NLMEMs.

It is well known that the sampling times in a PK study can have a large impact on the precision and bias of parameter estimates (D'Argenio (1981)). However, there is contention in the literature as to whether it is better to sparsely sample a larger number of individuals, or to heavily sample a smaller number of individuals in population studies. Sheiner and Beal (1983), Hashimoto and Sheiner (1991), and Jonsson et al. (1996) evaluate the effect of altering the number of subjects, and the number and timing of blood samples in PK studies on the precision and bias of the estimated parameter values. Sheiner and Beal (1983), and Hashimoto and Sheiner (1991) recommend that designs which use more study subjects, even if the majority are sparsely sampled, are preferable over designs which use more sampling times on fewer individuals. Conversely, Jonsson et al. (1996) recommend increasing the number of samples per subject, even if the total number of study subjects is small, so that the parameter estimates are unbiased and precise. However, these studies only investigate the use of a small number of sampling times (up to three sampling times), which were often fixed in their values and did not optimise a utility function over the design space.

4.1.2 Bayesian Hierarchical Model Framework

In the Bayesian framework, mixed effects models are commonly constructed using hierarchical models. These models account for the different levels of variability within and between populations. The observable outcomes are modelled conditionally on certain parameters, which are themselves assigned a probability distribution in terms of other parameters which are known as *hyperparameters*.

Here we consider NLMEMs where the j-th observation of individual i, $Y_{i,j}$, is given by:

$$Y_{i,j} = f(\boldsymbol{\phi}_i, \mathbf{d}_{i,j}) + \epsilon_{i,j},$$

where $f(\cdot)$ is a nonlinear mean function (which is the same for all individuals), $\boldsymbol{\phi}_i$ is the model parameter for individual i (i.e., the random effect), $\mathbf{d}_{i,j}$ is the experimental setting, and the errors are independent and distributed $\epsilon_{i,j} \sim N(0, \sigma^2)$. It is assumed that there are n subjects involved in the study.

A population distribution is specified for the individual parameter vectors $\boldsymbol{\phi}_i$, $i = 1, ..., n$:

$$\boldsymbol{\phi}_i \sim \text{MVN}(\boldsymbol{\lambda}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\Omega}$ are the population mean and variance-covariance matrix respectively. In this work we assume $\boldsymbol{\Omega}$ is known. Racine-Poon (1985) provides guidance on prior elicitation for unknown $\boldsymbol{\Omega}$. The population parameter (i.e., fixed effect) is denoted by $\boldsymbol{\phi}$. We define $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\phi}_{1:n})$, where $\boldsymbol{\phi}_{1:n} = (\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n)$. A log normal or inverse gamma prior (with shape hyperparameters a_0 and b_0) may be used for the observational variance, σ^2. Priors for the population parameters may also be specified as:

$$\boldsymbol{\phi} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

where $\boldsymbol{\mu}$ is the prior mean for $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ is the prior variance-covariance for $\boldsymbol{\phi}$. We specify values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in this work, based on simulation studies.

4.1.3 Bayesian Optimal Design Theory

The Bayesian optimal design framework involves defining a prior distribution for the population parameter $\boldsymbol{\phi}$ and a population model for the subject-specific parameter $\boldsymbol{\phi}_{1:n}$; a conditional sampling distribution $p(\mathbf{y}_i|\mathbf{d}, \boldsymbol{\phi}_i)$ for observing a new set of measurements $\mathbf{y}_i$ for individual i at the design points $\mathbf{d}$, given parameter values $\boldsymbol{\phi}_i$; and a utility function $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ that describes the reward that is obtained for taking measurements at design points $\mathbf{d}$, and observing the data $\mathbf{y}$ (where $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$), assuming the model parameter values $\boldsymbol{\theta}$ are known. In this work, we assume that measurements are taken at the same design points $\mathbf{d}$ for all study subjects.

The optimal Bayesian design, $\mathbf{d}^*$, maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$ with respect to the future data $\mathbf{y}$ and model parameter $\boldsymbol{\theta}$:

$$\begin{aligned}\mathbf{d}^* &= \arg\max_{\mathbf{d}\in\mathbf{D}} \; E\{U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})\} \\ &= \arg\max_{\mathbf{d}\in\mathbf{D}} \int_{\mathbf{Y}} \int_{\boldsymbol{\Phi}} \int_{\boldsymbol{\Phi}_{1:n}} U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \prod_{i=1}^{n} \Big\{ p(\mathbf{y}_i|\mathbf{d}, \boldsymbol{\phi}_i) p(\boldsymbol{\phi}_i|\boldsymbol{\phi}) d\boldsymbol{\phi}_i \Big\} p(\boldsymbol{\phi}) d\boldsymbol{\phi} d\mathbf{y}, \qquad (4.1)\end{aligned}$$

where $p(\mathbf{y}_i|\mathbf{d}, \boldsymbol{\phi}_i)$ is the conditional likelihood, $p(\boldsymbol{\phi}_i|\boldsymbol{\phi})$ is the population distribution for the random effects and $p(\boldsymbol{\phi})$ is the prior distribution for the fixed effects.

4.1.4 Likelihood Function Approximation for NLMEMs

Due to the nonlinearity, NLMEMs have no analytical expression for the observed data likelihood. The random effects must be integrated out from equation (4.1). There is a wealth of literature on approaches for approximating the observed data likelihood to perform inference for the population parameters. These methods include: first-order approximations (Sheiner and Beal (1983)); first order conditional methods (Lindstrom and Bates (1990)); Gaussian quadrature (e.g., Davidian and Gallant (1993)); adaptive Gaussian quadrature (e.g., Rabe-Hesketh et al. (2004)); Laplace approximations (e.g., Beal and Sheiner (2002));

Markov chain Monte Carlo (MCMC) (e.g., Spiegelhalter et al. (1996)); Monte Carlo integration (e.g., Wakefield (1994)); and importance sampling (Geweke (1989)). From the Bayesian perspective, the parameters $\boldsymbol{\phi}_{1:n}$ and $\boldsymbol{\phi}$ have the same status, i.e., inference for either $\boldsymbol{\phi}_{1:n}$ or $\boldsymbol{\phi}$ are made using the posterior distribution. This is not the case for frequentist inference where $\boldsymbol{\phi}_{1:n}$ are random variables and $\boldsymbol{\phi}$ is a constant.

4.1.5 Bayesian Designs for NLMEMs

Classical design criteria often consist of scalar functions of the Fisher information matrix (FIM), such as the determinant or the trace (e.g., Mentré et al. (1997); Retout and Mentré (2003)). Pseudo-Bayesian design criteria also consist of functions of the FIM, but also average these functions over a "prior" for the model parameters to account for parameter uncertainty (e.g., Pronzato and Walter (1985)). Once an expression for the likelihood has been found one can then derive the FIM.

Bayesian design criteria are often based upon the expected gain in Shannon information from the prior to posterior distribution (also known as 'mutual information' or the 'Kullback-Leibler distance') (e.g., Chaloner and Verdinelli (1995)). Other commonly-used Bayesian design criteria are based on the spread of the posterior distribution, which may be measured, for example, by the precision or entropy (e.g., Stroud et al. (2001)). When the posterior distribution is found by simulation, it must be sampled from for each future data set that is drawn from the prior predictive distribution, and so many thousands of posterior distributions are often required to perform Bayesian experimental design. For this reason, fully Bayesian experimental designs for NLMEMs are largely unexplored.

Han and Chaloner (2004) searched for Bayesian population designs for a HIV dynamics study. They did not optimise over a continuous design space, but instead considered 8 fixed sampling schedules. The posterior predictive variance for two parameters was used as the utility function, and MCMC was used to sample the posterior distribution for each future dataset. Palmer and Müller (1998) implemented Bayesian optimal designs for population models for determining the timing of stem cell collections in cancer patients. The NLMEMs were estimated by MCMC simulation and a discrete set of designs were searched over.

Stroud et al. (2001) fit NLMEMs to existing data relating to the PK of the anticancer agent, paclitaxel, in patients and found (sequential) Bayesian designs for the subject-specific parameters for the next patient. The design variable was the blood sampling times and two utility functions were used: the posterior precision of the area under the curve and the posterior precision of the time above a certain drug concentration. Each of these utility functions also included a cost penalty (which penalised sampling times that occurred after some pre-specified time) and were estimated using importance sampling. A Metropolis-Hastings MCMC algorithm (e.g., Müller (1999)) was used to find the optimal design.

4.1.6 Contribution and Outline

In this paper, we present static, fully Bayesian designs for population parameters of NLMEMs, in which we use simulation-based optimal design methods to search over both continuous and discrete design spaces. In this paper, we define a 'continuous design space' to be one in which the designs can take on any value in a pre-defined continuous interval, rather than values from a fixed set of discrete values. Whilst previous studies have found optimal Bayesian designs for NLMEMs by searching over a finite set of designs (e.g., Han and Chaloner (2004); Palmer and Müller (1998)), to our knowledge, no studies have searched over a continuous design space to find optimal static Bayesian designs for NLMEMs. This work is motivated by a PK study conducted by McGree et al. (2012d) which models the PK of an acepromazine metabolite in racing horses. Our design problem consists of finding the optimal sampling times for the PK study, as well as the optimal number of subjects to incorporate into the study, and the optimal number of samples to take per subject. To our knowledge, no previous Bayesian experimental design studies have addressed all of these issues (over a continuous design space). We are interested in finding designs that maximise the posterior precision of the population parameters, subject to cost constraints.

In Section 4.2 we describe the design methodology used in this work. In Section 4.3 we obtain some results for a simple example that involves designing for a linear mixed effects regression model. In Section 4.4 our PK case study is introduced and our design methods are applied to the case study in Section 4.5. The article concludes with a discussion in Section 4.6.

4.2 Bayesian Experimental Design Framework

Equation (4.1) does not usually have a closed form solution, and so numerical approximations or simulation methods are used to solve the maximisation and integration problem. These include: numerical quadrature or Laplace approximations (Brockwell and Kadane (2003); Ryan et al. (2014a)); prior simulation (Müller (1999)); MCMC simulation in an augmented probability model (Müller (1999)); and sequential Monte Carlo (Amzal et al. (2006)).

In this work, we use the approach implemented by Müller (1999) to solve equation (4.1). This involves the use of MCMC which samples from the target distribution:

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}),$$

using a Metropolis-Hastings scheme. $h(\cdot)$ is constructed in such a way that the marginal distribution $h(\mathbf{d})$ is proportional to the expected utility, $U(\mathbf{d})$. It is assumed that the utility $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ is non-negative and bounded, and that $h(\cdot)$ is integrable and can be normalised. The sample of simulated $\mathbf{d}$ may be used to provide an estimate of $h(\mathbf{d})$ and the joint mode of $h(\mathbf{d})$, $\mathbf{d}^*$, corresponds to the optimal design. The Metropolis-Hastings scheme of Müller (1999) is described in Algorithm 4.1 and has been adapted for designing for NLMEMs.

Algorithm 4.1: MCMC algorithm for Bayesian optimal design for NLMEMs

1 Set an initial design $\mathbf{d}^{(1)}$.

2 Draw $\boldsymbol{\phi} \sim p(\boldsymbol{\phi})$, $\boldsymbol{\phi}_i \sim p(\boldsymbol{\phi}_i|\boldsymbol{\phi})$, $\mathbf{y}_i \sim p(\mathbf{y}_i|\mathbf{d}^{(1)}, \boldsymbol{\phi}_i)$, for $i = 1, ..., n$ individuals.

3 Compute $U^{(1)} = U(\mathbf{d}^{(1)}, \boldsymbol{\theta}, \mathbf{y})$, where $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_n)$ and $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\phi}_{1:n})$.

4 **for** *j = 1 to iters* **do**

5 Generate a candidate design $\tilde{\mathbf{d}}$ from a proposal distribution $q(\cdot|\mathbf{d}^j)$.

6 Generate proposals for the parameters and simulate data:

7 $\tilde{\boldsymbol{\phi}} \sim p(\boldsymbol{\phi})$, $\tilde{\boldsymbol{\phi}}_i \sim p(\boldsymbol{\phi}_i|\tilde{\boldsymbol{\phi}})$, $\tilde{\mathbf{y}}_i \sim p(\mathbf{y}_i|\tilde{\mathbf{d}}, \tilde{\boldsymbol{\phi}}_i)$, for $i = 1, ..., n$ individuals.

8 Compute $\tilde{U} = U(\tilde{\mathbf{d}}, \tilde{\boldsymbol{\theta}}, \tilde{\mathbf{y}})$, where $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}_1, ..., \tilde{\mathbf{y}}_n)$ and $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\phi}}_{1:n})$.

9 Calculate the MH acceptance probability, $a = \min(1, A)$ where

$$A = \frac{\tilde{U} \times q(\mathbf{d}^{(j)}|\tilde{\mathbf{d}})}{U^{(j)} \times q(\tilde{\mathbf{d}}|\mathbf{d}^{(j)})}.$$

Here $U^{(j)}$ and $\mathbf{d}^{(j)}$ are the current utility and design point values, respectively, and $\tilde{U}$ and $\tilde{\mathbf{d}}$ are the proposed utility and design point values, respectively.

10 Set

$$(\mathbf{d}^{(j+1)}, U^{(j+1)}) = (\tilde{\mathbf{d}}, \tilde{U})$$

with probability a, and

$$(\mathbf{d}^{(j+1)}, U^{(j+1)}) = (\mathbf{d}^{(j)}, U^{(j)})$$

with probability $1 - a$.

11 **end for**

Simulation-based algorithms such as those presented by Müller (1999) have been found to have slow convergence for situations where there are a large number (≥ 4) of design variables (e.g., Stroud et al. (2001); Amzal et al. (2006)). For our PK application of interest, we are interested in searching for up to 15 sampling times. To ease the computational burden of having to search for a large number of design points, we use a lower dimensional parameterisation that reduces the design problem to one that involves searching over two design variables. The lower dimensional parameterisation was used in Line 5 in Algorithm 4.1. The sampling times for the PK study (Sections 4.4 and 4.5) will be generated from the evenly spaced percentiles of a Beta(a, b) distribution (see Ryan et al. (2014c)), scaled to $[0, 48]$ hours, where $a, b > 0$. Using this lower dimensional parameterisation, the Müller (1999) algorithm searches over the two design variables (a, b), and once these optimal values are found, a large number of design points can be generated from the evenly-spaced percentiles of the Beta(a, b) distribution. However, it must be noted that the designs generated by this lower dimensional parameterisation are not optimal but *near* optimal, which is a compromise of the computational savings achieved through these methods. We chose this lower dimensional parameterisation as we have used it previously for designing for fixed effects PK models (see Ryan et al. (2014c)) and found that it gave fairly flexible designs that could be suitable for PK studies.

The convergence of the MCMC algorithm was carefully monitored (through examination of autocorrelation plots, histograms and contour plots of the design variables, and trace plots of the utility functions over the iterations). To determine the optimal designs, we searched for the multivariate mode of the multivariate normal kernel smoothing density estimates of the design variables (see Cook et al. (2008); Drovandi and Pettitt (2013)).

4.2.1 Utility Function Estimation via Importance Sampling

Utility functions are problem-specific and incorporate the aims of an experiment. Bayesian utility functions are often based on the posterior distribution $p(\boldsymbol{\theta}|\mathbf{d},\mathbf{y})$. However, for NLMEMs, the posterior does not have a closed form expression and numerical methods are required for its approximation. We use a similar approach to Stroud et al. (2001) to generate samples from the posterior distribution, via importance sampling, for use in the Bayesian utilities.

Importance sampling is a commonly-used approach for approximating target distributions (Geweke (1989)) (in this case, the posterior $p(\boldsymbol{\phi}_{1:n},\boldsymbol{\phi}|\mathbf{d},\mathbf{y})$). It involves choosing an *importance* distribution $g(\cdot)$ from which it is easy to sample, and then weighting the samples to account for any differences between the importance and target distribution. The target and importance distributions should have the same support. Weighted discrete approximations, $\{(\boldsymbol{\phi},\boldsymbol{\phi}_{1:n})_j, W_j\}_{j=1}^{M_p}$, are produced (where M_p is the number of values or particles used) from the target distribution, where

$$w(\boldsymbol{\phi},\boldsymbol{\phi}_{1:n}) = \frac{\left\{\prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{d},\boldsymbol{\phi}_i)p(\boldsymbol{\phi}_i|\boldsymbol{\phi})\right\}p(\boldsymbol{\phi})}{g(\boldsymbol{\phi},\boldsymbol{\phi}_{1:n})}$$

are the importance weights, and $W_j \propto w((\boldsymbol{\phi},\boldsymbol{\phi}_{1:n})_j)$ are the normalised importance weights, $\sum_{j=1}^{M_p} W_j = 1$.

The prior distribution is commonly used as the importance distribution in Bayesian experimental design (e.g., Ryan et al. (2014c)), which reduces the importance weights to the (conditional) likelihood function. We use the population distribution (for the random effects) and the prior (for the fixed effects) as the importance distribution ($g(\boldsymbol{\phi},\boldsymbol{\phi}_{1:n}) = \left\{\prod_{i=1}^{n} p(\boldsymbol{\phi}_i|\boldsymbol{\phi})\right\}p(\boldsymbol{\phi})$), which reduces importance weights to the conditional likelihood:

$$w(\boldsymbol{\phi},\boldsymbol{\phi}_{1:n}) = \prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{d},\boldsymbol{\phi}_i).$$

This is not to be confused with the marginal likelihood $p(\mathbf{y}_{1:n}|\mathbf{d},\boldsymbol{\phi})$ in which the random effects have been integrated out. Note that the prior $p(\boldsymbol{\phi})$ for the fixed effects is relatively informative for our application as it is based on the results from the analysis of previous experiments.

To measure the efficiency of importance sampling, the effective sample size (ESS) is used, where

$$ESS = \frac{1}{\sum W_j^2}, 1 \leq ESS \leq M_p.$$

In this work we are only interested in the posterior distribution for the population parameters, and so we only use the samples of the subject-specific effects $\boldsymbol{\phi}_{1:n}$ to calculate the importance weights and discard them thereafter. Our method for approximating the

posterior distributions for the population parameters $p(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y})$ is outlined in Algorithm 4.2.

Algorithm 4.2: Algorithm for approximating $p(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y})$

1 Draw $\{\boldsymbol{\phi}^j\}_{j=1}^{M_p}$ from $p(\boldsymbol{\phi})$.
2 **for** $i = 1 : n$ **do**
3 Draw $\{\boldsymbol{\phi}_i^j\}_{j=1}^{M_p}$ from $p(\boldsymbol{\phi}_i|\boldsymbol{\phi}^j), j = 1, ..., M_p$
4 **end for**
5 $\{\boldsymbol{\phi}^j\}_{j=1}^{M_p}$ and $\{\boldsymbol{\phi}_{1:n}^j\}_{j=1}^{M_p}$ are only drawn once at the beginning of Algorithm 4.1 (prior to line 1) and are stored.
6 Weight $w^j = \prod_{i=1}^n p(\tilde{\mathbf{y}}_i|\tilde{\mathbf{d}}, \boldsymbol{\phi}_i^j), j = 1, ..., M_p$.
7 Normalise w^j to give $W^j, j = 1, ..., M_p$.
8 The particle approximation to $p(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y})$ is given by $\{\boldsymbol{\phi}^j, W^j\}_{j=1}^{M_p}$.

The utility functions can then be estimated using the weighted samples. Algorithm 4.2 (lines 6 - 8) is used in each iteration of Algorithm 4.1 (line 8, as well as line 3 at the beginning of the algorithm) to calculate the utility function. For our applications, we will use the determinant of the posterior precision matrix of the population parameters as the utility function:

$$U(\mathbf{d}, \mathbf{y}) = \det(\text{prec}(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y})).$$

For the example considered in Section 4.5, we set $M_p = 100000$ as this number of particles provided reasonably stable (based on the effective sample size) and precise estimates of the utility function.

4.3 Simple Illustrative Example: Linear Mixed Effects Model

We will begin with a simple example, in which we determine the optimal balance between the number of subjects (n) and the number of samples per subject (n_d), as well as the optimal values for the predictor variable $\mathbf{x}$, under certain conditions. The model is a linear mixed effects model where the response for the i-th subject is given by:

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\phi} + \mathbf{Z}_i\boldsymbol{\phi}_i + \boldsymbol{\epsilon}_i,\ i = 1, ..., n.$$

The observation vector $\mathbf{y}$ is of dimension $N \times 1$ (where N is the total number of observations). $\mathbf{X}$ is the design matrix for the fixed effects (of dimension $N \times p$) and $\mathbf{Z}_i$ is the design matrix for the i-th subject (of dimension $n_d \times p$). $\boldsymbol{\phi}$ consists of the $(p \times 1)$ fixed effects, with the prior $\boldsymbol{\phi} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a known, $p \times p$ nonsingular matrix. $\boldsymbol{\phi}_{1:n}$ consists of the $(np \times 1)$ random effects, with the subject-specific model:

$$(\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_n)^T \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\Omega}$ is a known, $np \times np$ nonsingular matrix. It is assumed that the random effects are independent. The residuals, $\boldsymbol{\epsilon}_i$, are independently distributed with $\boldsymbol{\epsilon}_i \sim \text{MVN}(\mathbf{0}, \mathbf{I}\sigma^2)$, $i = 1, ..., n$. Here we assume that σ^2 is known. It is also assumed that the observational errors ϵ_i are independent of the fixed or random effects.

For the linear mixed effects model, Sorensen and Gianola (2002) have derived an analytical expression for the posterior density of the fixed effects and we will use their expression for the posterior precision matrix to construct our Bayesian utility function. Since the integrals in (4.1) can be computed analytically, all that is required is to perform the optimisation to find the design which maximises the utility function.

We use a linear regression model of the following form:

$$y_{i,j} = (\phi_0 + \phi_{i,0}) + (\phi_1 + \phi_{i,1})x_{ij} + (\phi_2 + \phi_{i,2})x_{ij}^2 + \epsilon_{i,j},$$

where $i = 1, ..., n, j = 1, ..., n_d$. $\boldsymbol{\phi}^T = (\phi_0, \phi_1, \phi_2)$ and $\boldsymbol{\phi}_i^T = (\phi_{i,0}, \phi_{i,1}, \phi_{i,2})$.

We set $\boldsymbol{\Sigma} = \text{diag}(0.4, 0.3, 0.2)$; $\boldsymbol{\Omega} = \text{diag}(0.2, 0.15, 0.12, 0.2, 0.15, 0.12, ..., 0.2, 0.15, 0.12)$, where the number of repeats of the variances $(0.2, 0.15, 0.12)$ for the three random effect parameters depends on the number of subjects, n; and $\sigma^2 = 0.01$.

The utility function is the (log) determinant of the posterior precision of the fixed effects:

$$U(\mathbf{d}, \mathbf{y}) = \log(\det(\text{prec}(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y})). \tag{4.2}$$

In addition to precisely estimating the (fixed effects) model parameters, our design objectives also included a cost constraint. Since the posterior precision (of the fixed effects) and cost penalty may not be on the same scale, equation (4.2) was used to find designs, subject to a certain fixed maximum cost (Stigler (1971)).

The cost function penalised for the number of subjects in the study, and the total number of measurements taken in the study:

$$C = c_{\text{sub}} \cdot n + c_{\text{sample}} \cdot n \cdot n_d, \tag{4.3}$$

where c_{sub} is the cost per subject, and was set to \$50, and c_{sample} is the cost per sample, and was set to \$10. These values were arbitrarily chosen for illustrative purposes. The cost function was used to determine the different combinations of the number of subjects and samples per subject that could be included in the study for a fixed total cost. The (log) determinant of the posterior precision was calculated for each of these combinations and comparisons were made to see which yielded the highest value of the utility function.

The expression for the posterior precision of $\boldsymbol{\phi}$ is given by:

$$\text{prec}(\boldsymbol{\phi}|\mathbf{d}, \mathbf{y}) = (\mathbf{X'V^{-1}X} + \boldsymbol{\Sigma}^{-1}\sigma^2)\sigma^{-2}, \tag{4.4}$$

where $\mathbf{V} = \mathbf{Z\Omega Z'}\sigma^{-2} + \mathbf{I}$ (Sorenson and Gianola (2002)). Here, $\mathbf{Z}$ denotes the design matrix for the random effects (of dimension $N \times np$). If $\boldsymbol{\Omega} = \mathbf{0}$, then $\mathbf{V} = \mathbf{I}$. Note that, although the posterior precision does not depend on $\mathbf{y}$ here, the posterior mean is dependent on $\mathbf{y}$. The designs $\mathbf{d}$ enter the utility function via the design matrices $\mathbf{X}$ and $\mathbf{Z}$, and therefore only the term $\mathbf{X'V^{-1}X}$ in the right hand side of equation (4.4).

No. Subjects	No. samples per subject	Total no. samples	Optimal exact design for $x^{\dagger}$, ξ	Utility function value $U(d)$
9	3	27	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 1 & 1 & 1 \end{Bmatrix}$	15.55
8	4	32	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 1 & 2 & 1 \end{Bmatrix}$	15.47
7	5	35	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 2 & 2 & 1 \end{Bmatrix}$	15.26
6	7	42	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 2 & 3 & 2 \end{Bmatrix}$	14.77
5	10	50	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 3 & 4 & 3 \end{Bmatrix}$	14.72
4	13	52	$\xi = \begin{Bmatrix} 0 & 0.5 & 1 \\ 4 & 6 & 3 \end{Bmatrix}$	14.27

Table 4.1: Optimal designs and utility function values for different combinations of the number of subjects and the number of samples per subject for a fixed cost of \$750. †Note that the weights for the exact design (row 2 of ξ) correspond to the number of replicates that are to be taken on each support point (row 1 of ξ) for 1 study subject. The sum of the weights is equal to the number of samples that are to be taken per subject.

4.3.1 Results

For a fixed cost of \$750, we investigated different combinations of the number of subjects and the number of samples per subject, to determine the optimal balance between these two quantities. We also searched over the design space for the predictor $\mathbf{x}$, whose values were restricted to occur between 0 and 1.

The optimal exact designs were found using the Müller (1999) algorithm with 10000 iterations. According to classical design theory, the D-optimal design for a quadratic equation with three unknown fixed effects parameters should have three support points - one at either end of the design space and one at the centre (e.g., Pukelsheim (1993); Tan and Berger (1999)). Therefore, we decided to search over three predictor design variables, (x_1, x_2, x_3), which were the three support points, and two weights/integers for an exact design (ω_1, ω_2), where $\omega_3 = n_d - (\omega_1 + \omega_2)$. The weights determine the number of replicates that are to be placed on each support point. Therefore, $\mathbf{d} = (x_1, x_2, x_3, \omega_1, \omega_2)$. We also searched over 4 predictor design variables, but found that one of these design points was a replicate of one of the three support points.

Since an analytical expression of the utility function, $U(\mathbf{d})$, was available, the mode could easily be found by choosing the sample with the highest $U(\mathbf{d})$ value. To simplify matters, it was assumed that all subjects had the same number of observations at the same values of the predictor variable $\mathbf{x}$. The results are summarised in Table 4.1.

From Table 4.1 it can be seen that the optimal support points for the model occur at the middle of the design space (0.5), and at either end (0 and 1). This is similar to results obtained in the classical design literature (e.g., Pukelsheim (1993)). Preference for the replicates was given to the centre (0.5) of the design space, followed by the start of the design space (0).

It can also be determined from Table 4.1 that, for this application of interest, it is preferable to take a smaller number of samples from a larger number of individuals (rather than heavily sample a smaller number of individuals). This is in agreement with Diggle et al. (1994) who note that for a uniform correlation structure (of the errors), the addition of one repeated measure within a subject conveys less information on the fixed effects than the addition of an independent measure of a new study subject.

Prior Sensitivity

We are now interested in investigating the effect of the prior distribution for the fixed effects and the model for the random effects on the optimal number of subjects and samples per subject. We will begin by varying the prior for the fixed effects $\boldsymbol{\phi} \sim N(0, c\boldsymbol{\Sigma})$, where we will use the values $c = 0.1, 1, 2, 10, 100$. As $c \to \infty$, we would obtain the same results as in a frequentist paradigm. The population model for the random effects is the same as above. The same combinations of the number of subjects and samples per subject were used as in Table 4.1, and we will use the optimal $\mathbf{x}$ values from this table.

From Figure 4.1, it can be seen that there does not appear to be any variation in the optimal number of samples to take per subject as the prior variance for the fixed effects changes. That is, the designs do not change with the prior variance for the fixed effects, but the values of the utility function decrease as the prior variance increases. It appears that it is more useful to take a smaller number of samples from a larger number of individuals to precisely estimate the fixed effects, regardless of how precise our a priori knowledge of these effects is, for a fixed (and somewhat precise) value for the population variance of the subject-specific effects. This makes sense: since we already have precise knowledge about the subjects and wish to precisely estimate the population parameters, then we should incorporate more subjects into the study (even if they only have a few samples taken), regardless of our level of knowledge of the population parameters. Also, for a precise amount of knowledge of the population parameters ($c = 0.1$) there is less variation in the utility values for the different combinations of the number of samples per subject and number of subjects.

It is important to remember, with respect to our utility function (equation (4.4)), that

$$\det(\mathbf{X'V^{-1}X} + \boldsymbol{\Sigma}^{-1}\sigma^2) \neq \det(\mathbf{X'V^{-1}X}) + \det(\boldsymbol{\Sigma}^{-1}\sigma^2).$$

This means that varying the prior variance of the fixed effects ($c\Sigma$) may slightly affect the design when the determinant of the posterior precision is taken, where a smaller value of c will have a larger impact on the design compared to a larger value of c. If the trace of the posterior precision matrix were taken instead of the determinant, then varying the prior of the fixed effects would have no impact on the design since the trace is a linear function.

Now we investigate how changing the population variance of the subject-specific effects alters the designs: $\boldsymbol{\phi}_i \sim N(0, k\boldsymbol{\Sigma})$, where $k = 0.005, 0.05, 0.5, 1, 2, 20$. For the fixed effects, we will set the value of c to 10000, to give prior variance $10000\boldsymbol{\Sigma}$, which essentially gives

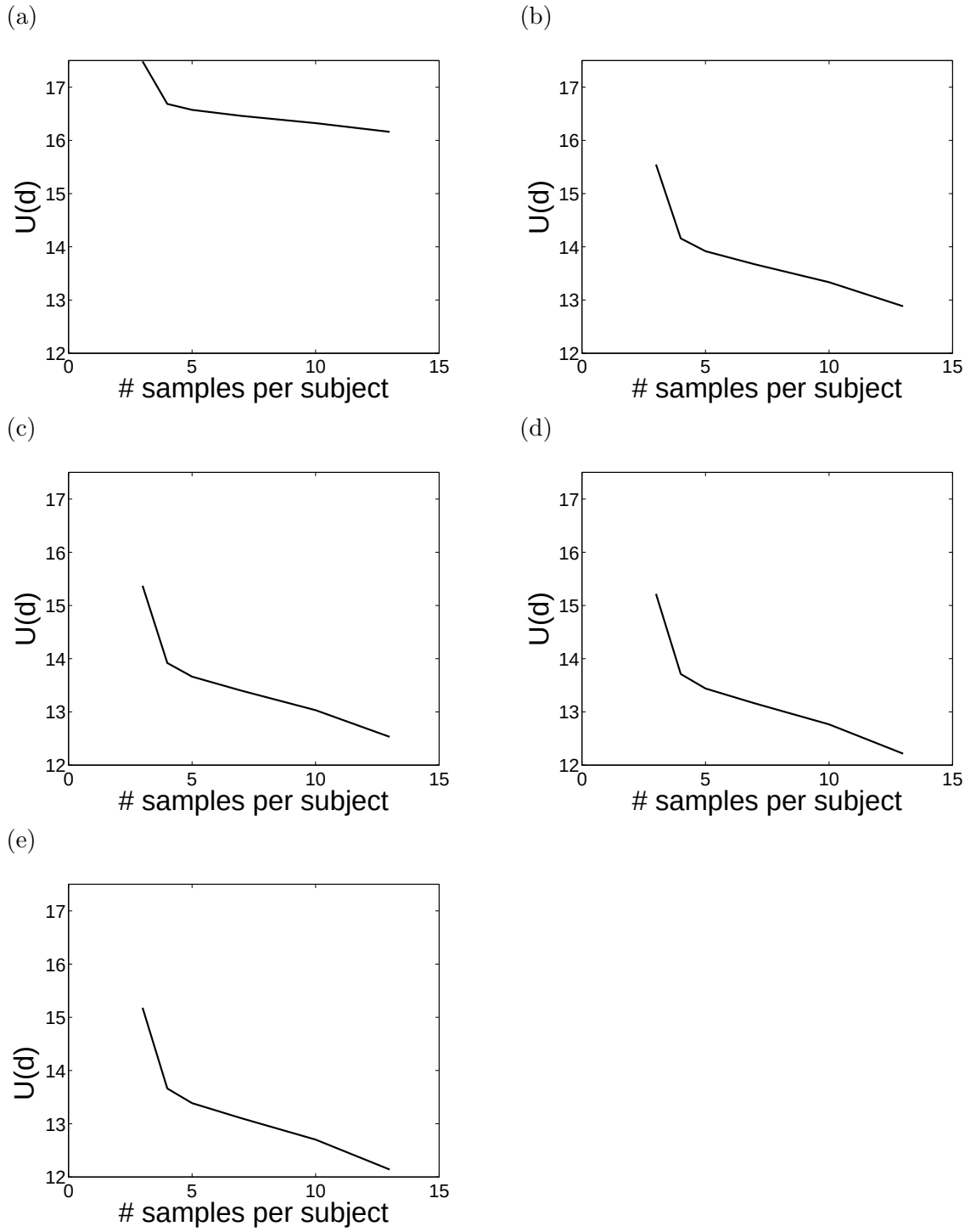


Figure 4.1: Utility function value versus the number of samples per subject, for a fixed cost of \$750, for various values of the prior variance for the population parameters for: (a) $c = 0.1$, (b) $c = 1$, (c) $c = 2$, (d) $c = 10$, and (e) $c = 100$.

a frequentist analysis. If $k \approx 0$, then $\text{Var}(\boldsymbol{\phi}_i) \approx 0$, and we will obtain the fixed effects frequentist result. We will investigate the same combinations of the number of subjects and samples per subject as above, and use the optimal $\mathbf{x}$ values from Table 4.1.

From Figure 4.2, it can be seen that as the population variance of the subject-specific effects increases, the optimal number of samples per subject decreases. That is, when there is a large amount of variation in the values of the subject-specific effects, one should focus on taking a small number of samples from a larger number of individuals, so that precise estimates of the population mean parameters can be obtained. If there is little variation in the subject-specific effects, then one should focus their resources on taking a larger number of samples from a smaller number of subjects, as there is little benefit from taking samples from more subjects to precisely estimate the population means.

An investigation into the sensitivity of the optimal design to the values chosen for the cost per subject and cost per sample was conducted and is presented in Appendix A. It is important to note that, although the same fixed total cost is obtained for the different combinations in Table 4.1, the total number of observations taken is not the same for each combination. In Appendix B, we assume that the same total number of observations is taken (and that the costs per subject and sample are equivalent), and investigate the 'best way' to divide up these observations.

4.4 Case Study: Population Pharmacokinetics of HEPS in Horses

This case study is concerned with determining the optimal urine sampling times for a population PK study of the acepromazine (ACP) metabolite 2-(1-hydroxyethyl)promazine (HEPS) in racing horses. We are also interested in determining the optimal number of horses to include in the study, and the optimal number of samples to take per horse (subject to cost constraints). Our case study will be a retrospective study design which makes use of the data collected and analysed by McGree et al. (2012d). We are interested in re-designing the study to precisely estimate the mean PK parameters of the population of horses.

In the study conducted by McGree et al. (2012d), 30mg of ACP was administered to 12 horses (geldings) and urine samples were taken at the following times: 2, 4, 6, 8, 12, 24, 36 and 48 hours after administration. The horses were trained to urinate to the sound of a whistle. Plasma samples were also taken but are not of interest to the current study. Here, we will re-design the urine sampling times from those used in McGree et al. (2012d) so that accurate measures of PK parameters of interest can be obtained.

4.4.1 The model

We assume n_d urine samples, $t_1, t_2, ..., t_{n_d}$, will be collected for n subjects. The cumulative amount of HEPS in subject i's urine at the j-th sampling time, y_{ij}, is modelled by:

$$Y_{ij} = f(\boldsymbol{\phi}_i, t_j) + \epsilon_{i,j},\ i = 1, ..., n; j = 1, ..., n_d,$$

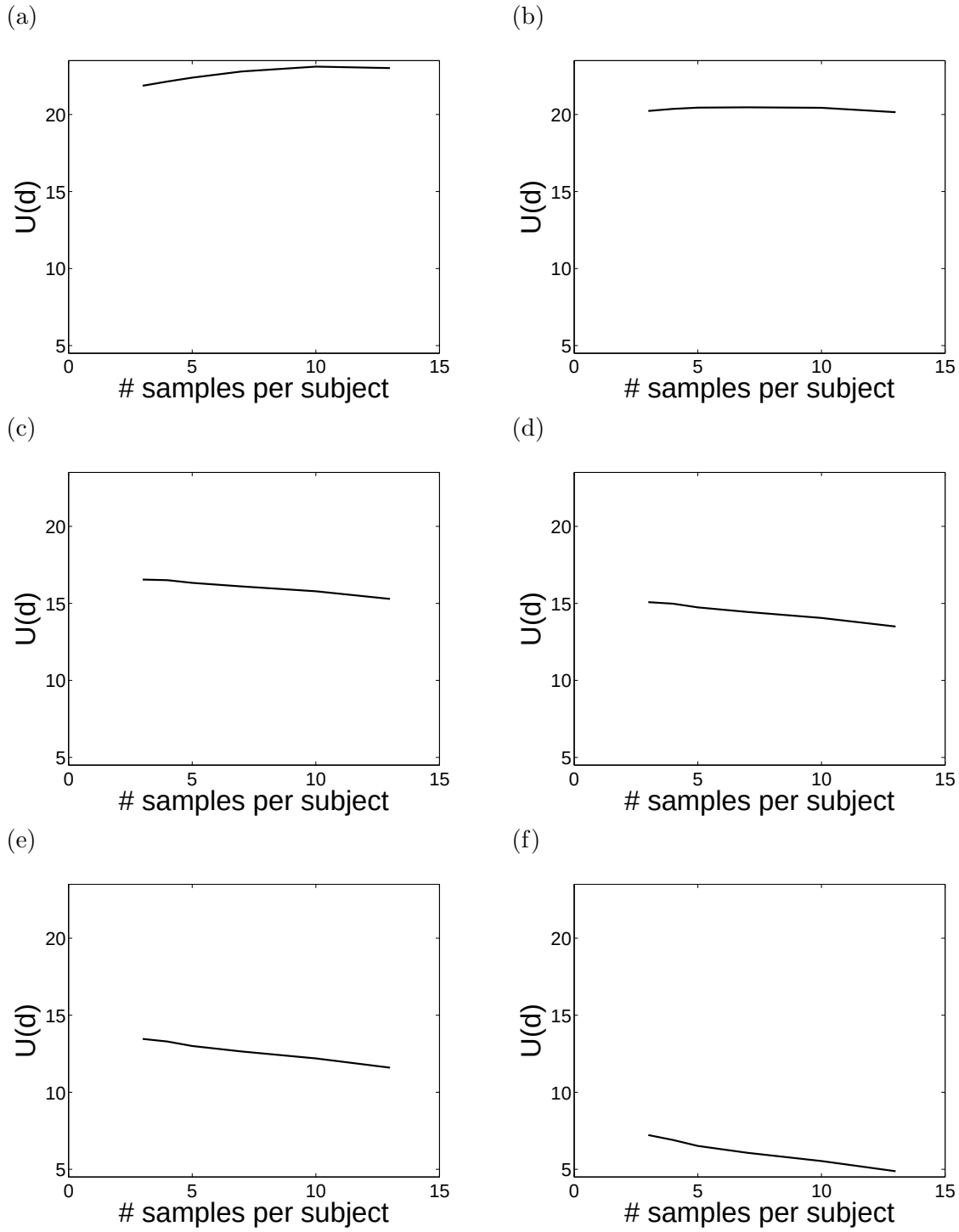


Figure 4.2: Utility function value versus the number of samples per subject, for a fixed cost of \$750, for various values of the variance for the subject-specific parameters: (a) $k = 0.005$, (b) $k = 0.05$, (c) $k = 0.5$, (d) $k = 1$, (e) $k = 2$, and (f) $k = 20$.

where

$$f(\boldsymbol{\phi}_i, t_j) = \frac{D \times F_{E_i}(Cl_i - V_i \times k_a - Cl_i \times e^{-k_a t_j} + V_i \times k_a \times e^{\frac{-Cl_i}{V_i} t_j})}{Cl_i - V_i \times k_a},$$

and independent $\epsilon_{i,j} \sim N(0, \sigma^2_{\text{add}})$.

Here $\boldsymbol{\phi}_i = (\log Cl_i, \log F_{E_i}, \log V_i)$ are the PK parameters for the i-th horse. Cl is the clearance rate, F_E is the fraction of HEPS that is excreted renally, and V is the volume of distribution. In this model it is assumed that $k_a = 35.87$ is a constant (as per McGree et al. (2012d)) and $D = 30000\mu$g is the drug dose. This model assumes a first-order absorption and elimination of the drug which is administered orally. The cumulative amount of the drug in the urine increases over time until all of the drug is eliminated.

Only additive error, whose variance is given by σ^2_{add}, is present in the model. It was assumed that $\sigma^2_{\text{add}} \sim N(1.2 \times 10^4, 3.1 \times 10^6)$ (based on McGree et al.'s (2012d) results).

The priors were obtained from the posterior results of McGree et al. (2012d), in which the Bayesian model above was fitted to the data using MCMC. The population distribution for the individual parameters $\boldsymbol{\phi}_i$, $i = 1, ..., n$, is specified as:

$$\boldsymbol{\phi}_i \sim \text{MVN}(\boldsymbol{\lambda}, \boldsymbol{\Omega}),$$

where $\boldsymbol{\lambda} = (\log Cl, \log F_E, \log V)$ are the population mean and $\boldsymbol{\Omega}$ is the population variance-covariance matrix. $\boldsymbol{\Omega}$ is assumed to be known and was obtained from the results of McGree et al. (2012d) and is given by:

$$\boldsymbol{\Omega} = \begin{pmatrix} 0.0149 & 0.0034 & -0.0037 \\ 0.0034 & 0.0146 & -0.0027 \\ -0.0037 & -0.0027 & 0.0048 \end{pmatrix}.$$

The fixed effects are assumed to have a prior distribution $\boldsymbol{\lambda} \sim \text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with known mean and variance-covariance matrix:

$$\boldsymbol{\mu} = \begin{pmatrix} 6.65 \\ -2.46 \\ 8.84 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 0.0076 & -0.0030 & 0.0050 \\ -0.0030 & 0.0050 & -0.0030 \\ 0.0050 & -0.0030 & 0.0073 \end{pmatrix}.$$

The estimated posterior densities of $\boldsymbol{\lambda} = (\log Cl, \log F_E, \log V)$ that were obtained by McGree et al. (2012d) are displayed in Appendix C. These posteriors were used as our prior for the retrospective design by fitting a multivariate normal distribution to the MCMC output of McGree et al. (2012d). Simulations from the prior predictive distribution for the cumulative urine amounts of HEPS are displayed in Figure 4.3.

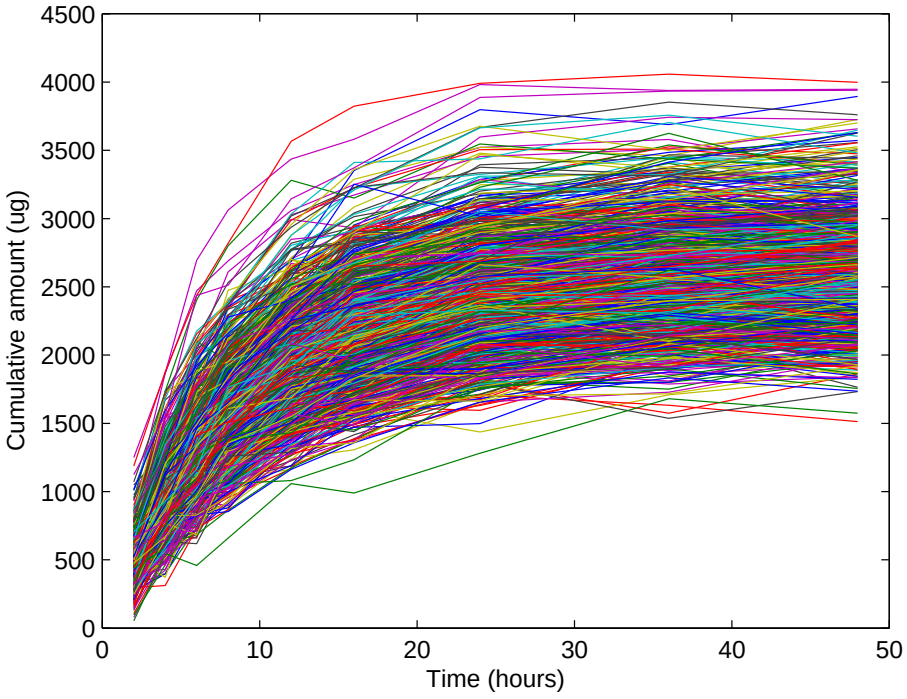


Figure 4.3: Prior predictive curves of cumulative urine amount of HEPS (online version in colour).

No. Subjects	No. samples per subject	Total no. of samples	Fixed cost ($)
2	10	20	$3000
3	5	15	
4	2	8	
3	11	33	$5000
4	7	28	
5	5	25	
6	3	18	
7	2	14	
5	15	75	$10 000
8	7	56	
12	3	36	

Table 4.2: Different combinations of the number of subjects and the number of samples per subject for various fixed total costs that were explored for the PK example.

4.5 Designs for Population PK Study

We now turn our attention to determining the optimal tradeoff between the number of subjects and samples per subject, and the optimal sampling times for a PK study. The methods discussed in Section 4.2 will be applied to the horse PK study introduced in Section 4.4. Table 4.2 displays the combinations of the number of subjects and samples per subject that are investigated for fixed total costs of $3000, $5000, and $10000.

The utility function is the determinant of the posterior precision of the population parameters and will be estimated using the procedure described in Section 4.2.1. We will implicitly use equation (4.3) as the cost function, where c_{sub} is the cost per horse, and was set to $500, and c_{sample} is the cost per sample, and was set to $100. These values

were determined after consulting several horse PK experts. Separate MCMCs were run for each of the combinations of the number of horses and number of samples per horse (i.e., separate MCMCs were run in parallel for each line of Table 4.2).

4.5.1 Results

We search for the (near) optimal urine sampling times, which are restricted to occur between 0 and 48 hours and use the lower dimensional parameterisation discussed in Section 4.2. It was assumed that all of the horses were sampled at the same times following the administration of the drug. The results of the sampling times that were generated from the (evenly-spaced) percentiles of a Beta(a, b) distribution for the different combinations of the number of subjects and samples per subject are displayed in Figure 4.4. The MCMC convergence diagnostics discussed in Section 4.2 were satisfied for all simulations that were performed, and Appendix D displays the convergence diagnostics for the MCMC simulations that were performed when $n = 3$ and $n_d = 11$ (i.e., the "3 horses, 11 sampling times per horse" combination).

For each of the total fixed costs considered, it was found that it was preferable to heavily sample a smaller number of subjects (see Figure 4.4). That is, if one is interested in obtaining precise posterior distributions of the population (urine) PK parameters, then one should heavily sample (10-15 samples per individual) a small number, say 2-5, of individuals. From Figure 4.4, it can be seen that the more subjects included and samples taken in the study the better, as expected, provided there is no upper limit to the cost of the study (which would be rare in practice). However, these designs may not be suitable in practice as it may be difficult or unethical to take a large number, say, 15 urine samples from an individual within 48 hours. It should be noted that these conclusions are subject to the cost ratios used in this study.

Our results are in agreement with Jonsson et al. (1996), but are in contrast to the results obtained by Sheiner and Beal (1983), and Hashimoto and Sheiner (1991), who recommend that designs that use a larger number of subjects that are sparsely sampled are preferable over designs that use a smaller number of individuals that are heavily sampled. However, their designs are not fully Bayesian and have different design objectives to those considered in this paper. Also, these studies only consider up to three sampling times, which are fixed in their values, and do not optimise the sampling times over the design space as we have done here. These results highlight the fact that the optimal sampling strategy may not be obvious and optimisation of the design problem is required using the methods we have described.

The (near) optimal sampling times were evenly spread across the design space (see Figure 4.4). For a small number of sampling times, preference was given to sampling times that spanned the central region of the design space (e.g., 10 - 40 hours, Figure 4.4). This is the region of the PK curve where the increase in the cumulative amount of HEPS in urine begins to progress at a slower rate and eventually asymptotes (see Figure 4.3). As the number of sampling times increased, they were evenly spread out from the central region of the design space to cover a greater region of the PK curve. The earliest and latest

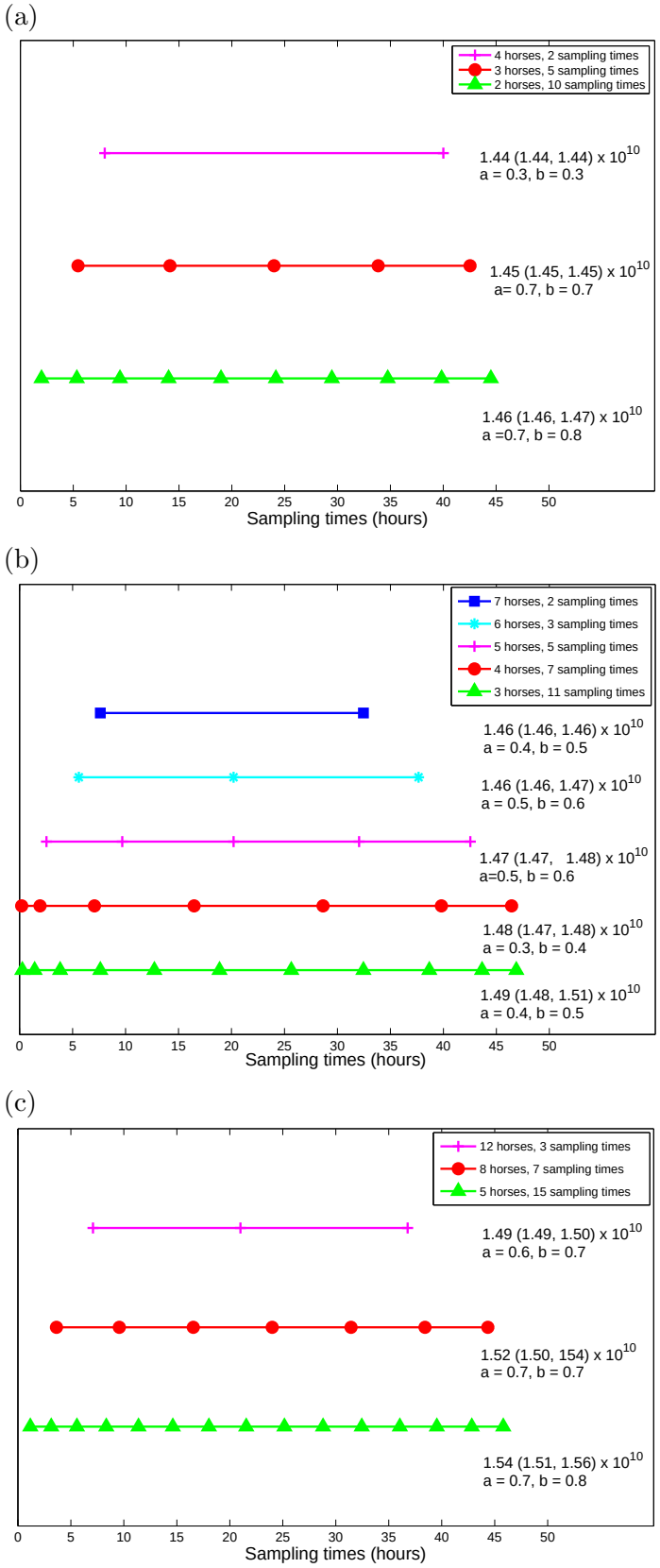


Figure 4.4: PK sampling times generated by the evenly spaced percentiles of a Beta(a,b) distribution, for the various combinations of the number of horses and number of sampling times per horse, for a fixed cost of (a) \$3000, (b) \$5000, and (c)\$10000. The utility function values are displayed next to the sampling times, along with the values for the shape parameters for the beta distribution that was used to generate the sampling times.

sampling times occurred at 0.12 and 47.4 hours respectively, and were associated with the 5 horses, 15 sampling times combination. These sampling times covered the majority of the design space [0, 48] hours. For the majority of the population designs (where more than 3 samples were taken), sampling continued after the cumulative amount of drug had reached an asymptote.

It should be noted that there was little difference in the utility function values for the different combinations of the number of subjects and samples per subject, for a fixed total cost. In Figure 4.4(a) the utility function values ranged from 1.44 - 1.46, in Figure 4.4(b) they ranged from 1.46 - 1.49, and ranged from 1.49 - 1.54 in Figure 4.4(c) (on average). Therefore it is difficult to make strong statements on the optimal sampling strategies, based on these results. Alternative priors for the fixed effects or models for the random effects may produce utility surfaces that are less flat.

We also tried implementing Algorithm 4.1 to search for the optimal number of subjects, number of samples per subject, and sampling times all at once, but this was found to be too computationally intensive.

A prior sensitivity study was also conducted (in a similar fashion to Section 4.3.1) and can be found in Appendix E. It was found that the designs did not vary when the prior variance for the population parameters was altered (similar to Figure 4.1), and it was most useful to heavily sample a smaller number of subjects. When the variation between the subjects (population variance) was small, it was preferable to take more samples from a smaller number of individuals, as there was little benefit from sampling a larger number of individuals. When the variation between the subjects was larger, it was preferable to take a smaller number of samples from more subjects, so that precise posterior distributions of the population parameters can be obtained. This is similar to Figure 4.2.

4.6 Discussion

In this article we have discussed and presented methods that can be used to find optimal fully Bayesian designs for mixed effects models, for both linear and non-linear models. The design problem was to determine the optimal number of subjects, the optimal number of samples per subject and the optimal predictor variable values (as in the linear model example) or the near optimal sampling times for a PK study (nonlinear example), to precisely estimate the model parameter of interest, subject to a cost constraint. Whilst the computational methods used in this work are not novel, their adaptation and application to find fully Bayesian static optimal designs for NLMEMs is new. Searches over a number of different design variables (some of which had a continuous design space) were performed, which also has not been previously implemented to find fully Bayesian static designs for NLMEMs.

Population designs comprise of a set of elementary designs that are to be carried out on groups of subjects. The elementary designs consist of several values of the design variable (e.g., blood sampling times, treatment doses etc) that are to be performed on each subject belonging to the design. The number of samples to be taken and the values of

the design variable may differ between subjects within an elementary design, and between the elementary designs. For simplicity, we assumed that all individuals in the examples considered in this paper had the same number of measurements taken and were sampled at the same experimental design. Previous simulation population PK studies (e.g., Sheiner and Beal (1983); Jonsson et al. (1996)) have found that the precision and accuracy of the parameter estimates are affected by the number of elementary designs, the number of subjects per elementary design, and the number and allocation of the design points (e.g., sampling times). Therefore, our design set up (in terms of having one elementary design) may not be optimal for population studies and future studies may wish to investigate the use of different elementary designs for different groups of subjects. This is likely to be very computationally intensive as many design variables would be involved.

In both the linear and nonlinear examples, we were also interested in determining the optimal number of subjects and samples per subject, subject to a cost constraint. This was achieved by searching over several different combinations of the number of subjects and samples per subject that resulted in the same total fixed cost. We had adapted the Müller (1999) algorithm to treat the number of subjects and the number of samples per subject as design variables, so that the optimal number of subjects, samples per subject and sampling times could be found simultaneously. However, one cannot search over a large number of design variables using the Müller (1999) algorithm as it becomes too computationally intensive to search over the joint space $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ and to determine the multivariate mode for the large number of design variables.

The results obtained in this study are dependent on the cost constraints and prior distributions used for the design problems, and for both examples, it is likely that different sampling strategies would have been achieved if different cost constraints or prior distributions were used. For the linear mixed effects model that was considered in this work, it was found that it was more useful to take a small number of samples from a larger number of individuals if one is interested in obtaining precise posterior distributions of the population parameters. For the NLMEM example (PK study), it was found that it was preferable to heavily sample a smaller number of individuals, so that precise posterior distributions of the population PK parameters could be obtained. The differences in the optimal sampling strategies between the two examples considered in this paper highlights how problem-dependent optimal Bayesian designs for mixed effects models is, emphasising the need for the optimisation methods presented in this paper (and references therein).

A lower dimensional parameterisation was used to reduce the computational burden of searching over a large number of design points. The MCMC algorithm searched over the two design variables (a, b), and once these optimal values are found, the design points were generated from the evenly-spaced percentiles of the Beta(a, b) distribution, scaled to $[0, 48]$ hours. We have previously found the beta proposal scheme to be quite flexible in generating designs, in that a wide variety of designs can be generated from this scheme depending on the values of the shape parameters used (Ryan et al. (2014c)). This parameterisation could be extended to offer further flexibility by including another design variable that determines the optimal percentiles of the beta distribution to use. It must

be stressed that the designs generated by these lower dimensional parameterisations are not optimal but near optimal.

In the examples considered in this paper, we were able to estimate the posterior density (for use in the Bayesian utility functions) analytically, or via importance sampling. Future studies that design for mixed effects models in a Bayesian framework should investigate alternative methods for estimating the posterior density, such as adaptive Gaussian quadrature (e.g., Rabe-Hesketh et al. (2004)) or Laplace approximations (e.g., Wolfinger (1993)), which may prove to be computationally faster and more efficient. We found that importance sampling from the prior (fixed effects) and population distribution (random effects) was somewhat computationally intensive as many importance samples ($M_p = 100000$) were required to obtain reasonably stable (based on the ESS) and precise estimates of the utility function.

Acknowledgments

E.G. Ryan was supported by an APA(I) Scholarship which came from an ARC Linkage Grant with Roche Palo Alto (LP0991602). The work of A.N. Pettitt was also supported by an ARC Discovery Project (DP110100159). The authors would like to thank Dr James McGree from Queensland University of Technology for his helpful comments throughout the process of writing this paper; for supplying the posterior samples for the horse PK model; and for consultation about his paper (McGree et al. (2012d)).

Supplementary Materials - Appendices

Appendix A - Cost Sensitivity Study for the Linear Mixed Effects Model Example (Section 4.3.1)

To investigate the sensitivity of the optimal design to the values chosen for the cost per subject and cost per sample, we varied these values from those used previously. A total fixed cost of \$750 was used, and we searched for the optimal number of subjects and samples per subject, as well as the optimal values of $\mathbf{x}$. The priors and model that were given at the beginning of Section 4.3 of the main paper were used. The results are summarised in Table 4.3.

Table 4.3 supports the results of Table 4.1 in the main paper, in that it is preferable to take a small number of samples from a larger number of subjects, for this linear mixed effects model and priors. The optimal designs are also similar to those in Table 4.1 (of the main paper), in that 3 support points were found at 0, 0.5 and 1, and that preference was given to the middle support point (0.5) followed by the first support point (0), as indicated by the weights.

Appendix B - No Cost Constraint for the Linear Mixed Effects Model Example (Section 4.3.1)

It is important to note that, although the same fixed total cost is obtained for the different combinations in Table 4.1 (of the main paper) and Table 4.3 (Appendix A), the total

Cost per subject	Cost per sample	No. Subjects	No. samples per subject	Total no. samples	Optimal exact design for x	Utility function value $U(d)$
\$30	\$20	7	3	21	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 1 & 1 & 1 \end{array} \right\}$	14.89
		6	4	24	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 1 & 2 & 1 \end{array} \right\}$	14.71
		5	6	30	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 2 & 3 & 1 \end{array} \right\}$	14.49
		4	7	28	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 2 & 3 & 2 \end{array} \right\}$	13.99
		3	11	33	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 3 & 5 & 3 \end{array} \right\}$	13.48
		2	17	34	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 5 & 8 & 4 \end{array} \right\}$	12.67
\$100	\$10	5	5	25	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 2 & 2 & 1 \end{array} \right\}$	14.38
		4	8	32	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 1 & 2 & 1 \end{array} \right\}$	14.06
		3	15	45	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 5 & 6 & 4 \end{array} \right\}$	13.58

Table 4.3: Optimal designs and utility function values for different combinations of the number of subjects and the number of samples per subject for different costs per subject and per sample, for a fixed total cost of \$750 for the linear mixed effects model.

No. Subjects	No. samples per subject	Optimal exact design for x	Utility function value $U(d)$
24	2	$\xi = \left\{ \begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right\}$	15.93
16	3	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 1 & 1 & 1 \end{array} \right\}$	17.12
12	4	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 1 & 2 & 1 \end{array} \right\}$	16.58
8	6	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 2 & 3 & 1 \end{array} \right\}$	15.73
6	8	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 2 & 4 & 2 \end{array} \right\}$	15.11
4	12	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 4 & 5 & 3 \end{array} \right\}$	14.23
3	16	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 5 & 7 & 4 \end{array} \right\}$	13.61
2	24	$\xi = \left\{ \begin{array}{ccc} 0 & 0.5 & 1 \\ 7 & 11 & 6 \end{array} \right\}$	12.75

Table 4.4: Optimal designs and utility function values for different combinations of the number of subjects and the number of samples per subject that result in a total number of 48 observations to be taken.

number of observations taken is not the same for each combination. We will now assume that 48 observations are taken in total and that the cost per individual and cost per sample is equivalent, and will investigate the 'best way' to divide up these observations, in terms of the number of subjects and samples per subjects. The results are summarised in Table 4.4.

From Table 4.4, it appears that the best way to divide up 48 observations is to take 3 samples each from 16 individuals. This makes sense, as we are interested in precisely

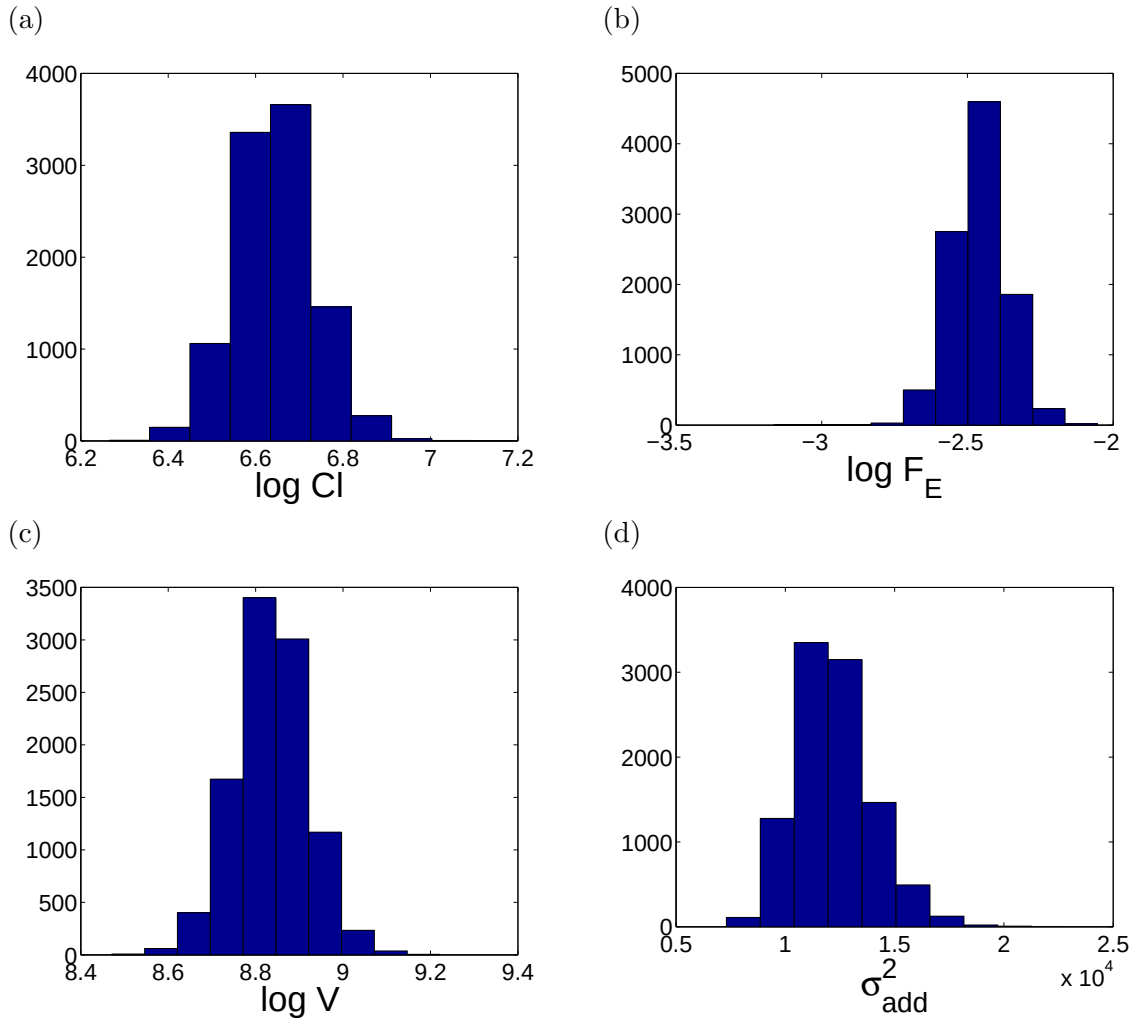


Figure 4.5: Estimated posterior densities of (a) $\log Cl$, (b) $\log F_E$, (c) $\log V$, and (d) σ^2_{add} from the MCMC simulations which fitted the model to the data for 12 horses.

estimating 3 population parameters, and so it is best to take 3 samples from a larger number (say, 16) of individuals to learn about the population parameters. These results are also in agreement with classical D-optimal designs in which p support points should be used and N/p (here $N = 48$) subjects should be used (Atkinson and Donev (1992); Tekle et al. (2007)).

Appendix C - Posterior Densities Obtained by McGree et al. (2012d) (Section 4.4)

The estimated posterior densities of $\boldsymbol{\phi} = (\log Cl, \log F_E, \log V)$ and σ^2_{add} that were obtained by McGree et al. (2012d) are displayed in Figure 4.5. These posteriors were used as our prior for the retrospective design by fitting a multivariate normal distribution to the MCMC output of McGree et al. (2012d).

Appendix D - MCMC Convergence Diagnostics (Section 4.5.1)

The convergence diagnostics for the MCMC simulations that were performed when $n = 3$ and $n_d = 11$ (i.e., the "3 horses, 11 sampling times per horse" combination) are displayed

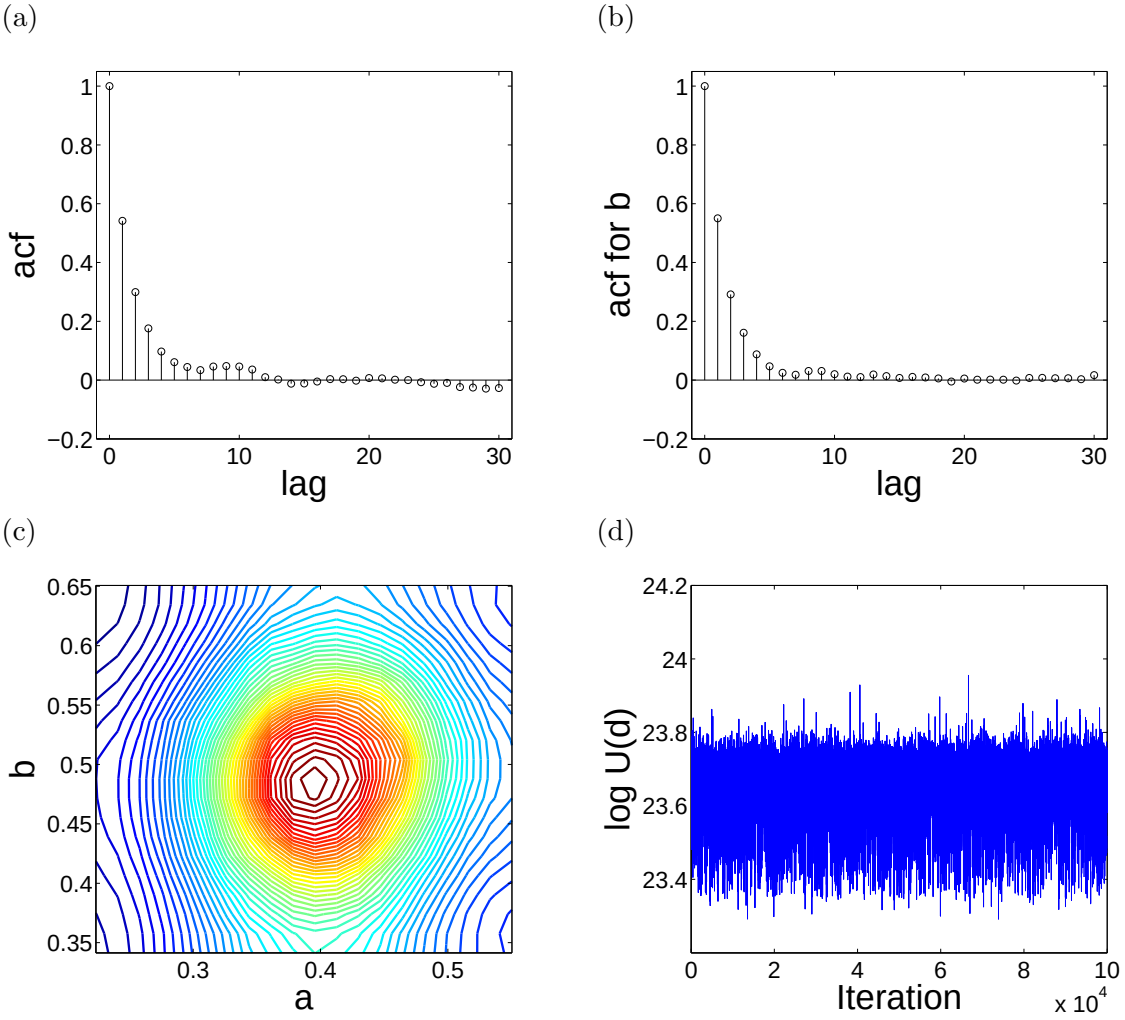


Figure 4.6: MCMC convergence diagnostics where $n = 3$ and $n_d = 11$: (a) autocorrelation plot for the design variable a, (b) autocorrelation plot for the design variable b, (c) contour plot for the design variables a and b, and (d) trace plot for the utility function.

in Figure 4.6.

From Figure 4.6, we can see that the MCMC simulations have converged.

Appendix E - Prior Sensitivity Study for PK Example (Section 4.5.1)

The sensitivity to the prior was investigated in a similar fashion to Section 4.3.1 of the main paper. The prior for the fixed effects was varied

$$\boldsymbol{\phi} \sim \text{MVN}(\boldsymbol{\mu}, c\boldsymbol{\Sigma}),$$

where we will use the values $c = 0.5, 1, 2$ and 5. The values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the same as those defined in Section 4.4 of the main paper, and the same model was used for the random effects as that which is given in Section 4.4 (of the main paper). The same combinations of the number of subjects and samples per subject were used as in Table 4.2 (of the main paper) for the fixed cost of \$5000, and we will use the optimal sampling

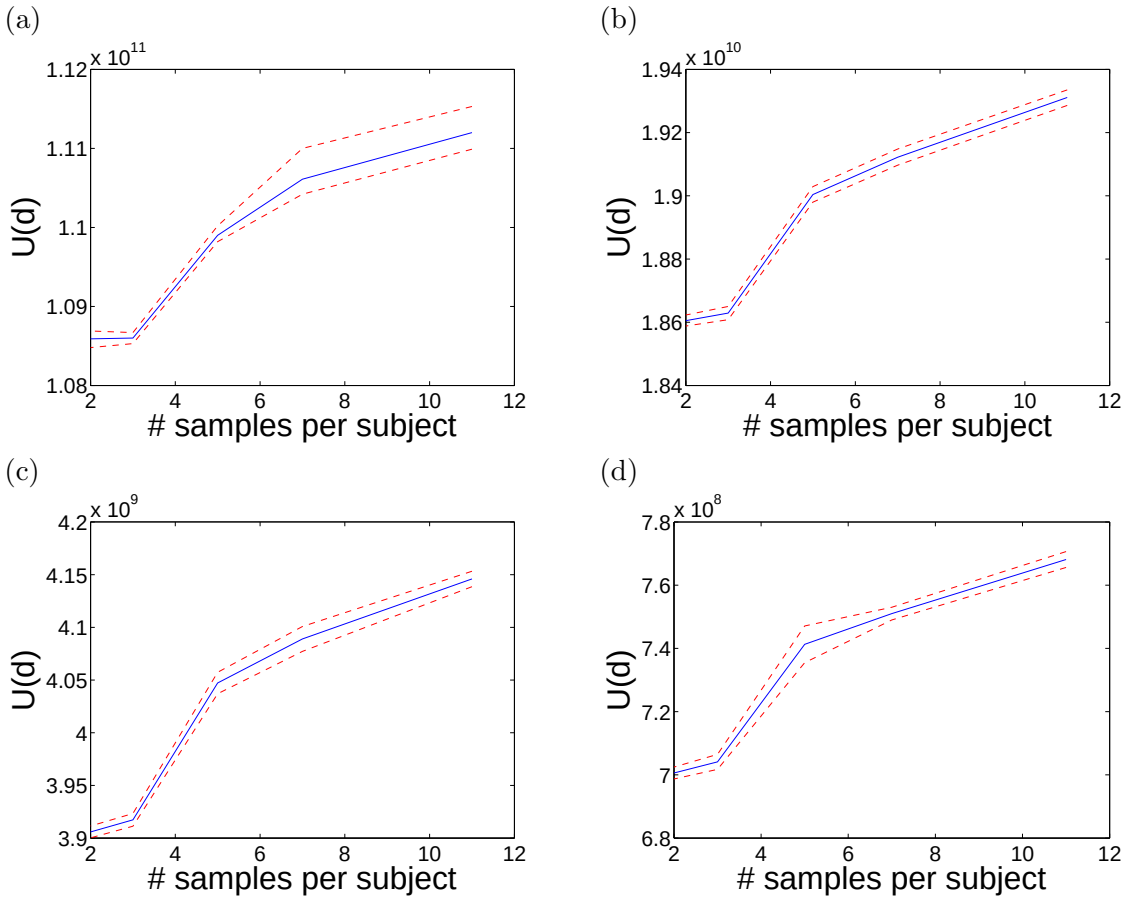


Figure 4.7: Utility function value versus the number of samples per subject, for a fixed cost of \$5000, for various values of the prior variance for the population parameters: (a) $c = 0.5$, (b) $c = 1$, (c) $c = 2$, and (d) $c = 5$. The solid line represents the utility function value and the dotted lines are a 95% confidence interval.

times from Figure 4.4 (b) (from the main paper). The results are summarised in Figure 4.7.

From Figure 4.7, it appears that there is no variation in the optimal number of samples per subject as the prior variance of the population parameters changes. This could be due to the fact that the priors investigated in Figure 4.7 are still quite informative. It appears that it is most useful to heavily sample a smaller number of subjects

There is variation in the utility function values where a higher utility function value was obtained when a smaller prior variance was used for the fixed effects. This is similar to Figure 4.1 (of the main paper). However, there is a computational efficiency effect present in Figure 4.7 that was not present in Figure 4.1 (of the main paper), due to the fact that importance sampling was used to estimate the posterior distribution for the NLMEM. As the value of c increases, the efficiency of importance sampling decreases and more particles are required to obtain reasonable estimates of the utility. Priors with a larger value of c (larger prior variance) than those presented here were also investigated, but resulted in poor estimates of the utility function (large standard errors), due to the poor performance of importance sampling when using a diffuse prior.

Now we will alter the value of the population variance of the subject-specific parameters (i.e., the random effects model) to investigate sensitivity to the model, using

$$\boldsymbol{\phi}_i \sim \text{MVN}(\boldsymbol{\lambda}, k\boldsymbol{\Omega})$$

where $k = 0.01, 0.1, 1, 2, 10,$ and 20. The prior for the fixed effects will be the same as in Section 4.4 of the main paper. The same combinations of the number of subjects and samples per subject were used as in Table 4.2 (of the main paper) for the fixed cost of \$5000, and we will use the optimal sampling times from Figure 4.4 (b) (of the main paper). The results are summarised in Figure 4.8.

From Figure 4.8 it can be seen that when there is a small amount of variation between the subjects, it is preferable to take more samples from a smaller number of individuals, as there is little benefit from sampling a larger number of individuals. When there is a larger amount of variation between the subjects, it is preferable to take a smaller number of samples from more subjects, so that precise posterior distributions of the population parameters can be obtained. This is similar to Figure 4.2 (of the main paper). However, the 95% confidence intervals are very large for $k = 10$ and $k = 20$, due to the fact that the importance sampling was not as efficient for these larger values of k, and the ranges of the mean utility function values vary substantially in Figure 4.8 (a) - (f).

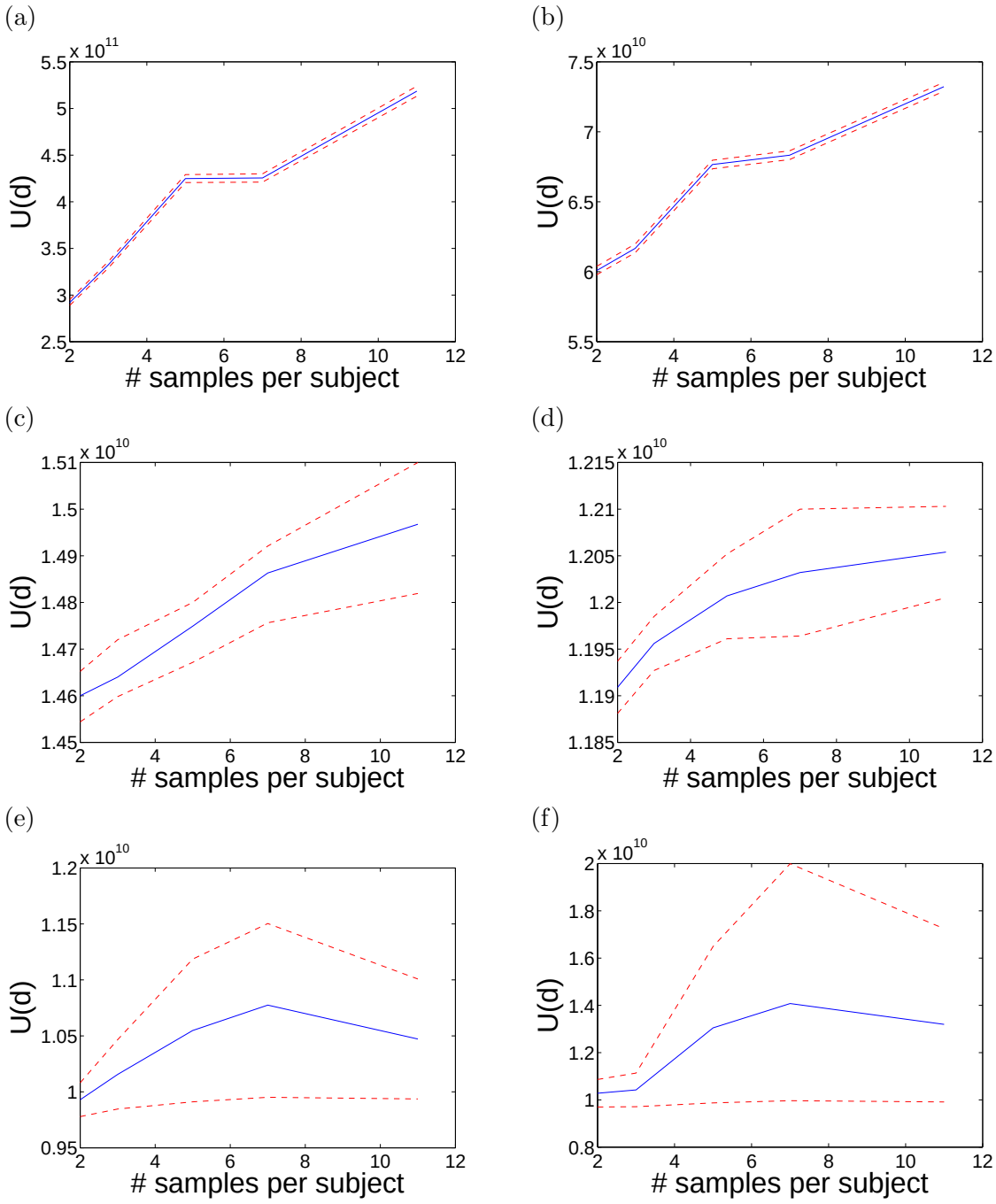


Figure 4.8: Utility function value versus the number of samples per subject, for a fixed cost of \$5000, for various values of the population variance for the subject-specific parameters: (a) $k = 0.01$, (b) $k = 0.1$, (c) $k = 1$, (d) $k = 2$, (e) $k = 10$, and (f) $k = 20$. The solid line represents the utility function value and the dotted lines are a 95% confidence interval.

Statement of Authorship for Chapter 5

This chapter has been written as a journal article. The authors listed below have certified that:

- They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
- They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
- There are no other authors of the publication according to these criteria;
- Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit, and
- They agree to the use of the publication in the students thesis and its publication on the QUT ePrints database consistent with any limitations set by publisher requirements.

In the case of this chapter, the reference for the associated publication is: **Ryan, E.G.**, Drovandi, C.C. and Pettitt, A.N. (2014). Fully Bayesian Experimental Design for Pharmacokinetic Studies. Submitted to *Entropy*.

Contributor	Statement of contribution
Elizabeth Ryan	Wrote all Matlab code required for the manuscript, performed all computations in the manuscript, interpreted and reported the results, constructed all figures presented in the manuscript, wrote the manuscript, and acted as the corresponding author
Signature and Date:	
Christopher Drovandi	Assissted in the writing of the Matlab code, directed the research and proofread the manuscript.
Tony Pettitt	Directed the research and proofread the manuscript.

Principal Supervisor Confirmation

I have sighted email or other correspondence from all co-authors confirming their certifying authorship.

Name	Signature	Date

Chapter 5

Fully Bayesian Experimental Design for Pharmacokinetic Studies

ABSTRACT

Utility functions in Bayesian experimental design are usually based on the posterior distribution. When the posterior is found by simulation, it must be sampled from for each future data set drawn from the prior predictive distribution. Many thousands of posterior distributions are often required. A popular technique in the Bayesian experimental design literature which rapidly obtains samples from the posterior is importance sampling, using the prior as the importance distribution. However, importance sampling from the prior will tend to break down if there is a reasonable number of experimental observations and/or the model parameter is high dimensional. In this paper we explore the use of Laplace approximations in the design setting to overcome this drawback. Furthermore, we consider using the Laplace approximation to form the importance distribution to obtain a more efficient importance distribution than the prior. The methodology is motivated by a pharmacokinetic study which investigates the effect of extracorporeal membrane oxygenation on the pharmacokinetics of antibiotics in sheep. The design problem is to find 10 near optimal plasma sampling times which produce precise estimates of pharmacokinetic model parameters/measures of interest. We consider several different utility functions of interest in these studies, which involve the posterior distribution of parameter functions.

KEYWORDS:
Bayesian design; Importance sampling; Laplace appromixation; Pharmacokinetics; Utility function

5.1 Introduction

5.1.1 Background

The selection of optimal conditions for conducting experiments is crucial to maximise the worth of data, especially for situations in which experiments are costly and/or time-consuming to conduct. Optimal experimental design aims to address these issues and may be employed to achieve the experimental goals in a more rapid and economical manner.

Bayesian methodologies for optimal experimental design have become more prominent in the literature (e.g., Müller (1999); Han and Chaloner (2004); Amzal et al. (2006); Müller et al. (2006); Cook et al. (2008); Huan and Marzouk (2013)). For an introduction

to Bayesian experimental design, see Chaloner and Verdinelli (1995). Bayesian optimal design involves defining a utility function $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ that describes the worth (based on the experimental aims) of choosing the design $\mathbf{d}$ from the design space $\mathbf{D}$ yielding data $\mathbf{y}$, with model parameter value $\boldsymbol{\theta}$. For example, the utility function could be the posterior precision of some parameter of interest. A probabilistic model, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$, is also required. This consists of a likelihood $p(\mathbf{y}|\mathbf{d}, \boldsymbol{\theta})$ for observing a new set of measurements $\mathbf{y}$ at the design points $\mathbf{d}$, given parameter value $\boldsymbol{\theta}$, and a prior distribution $p(\boldsymbol{\theta})$ for the parameter $\boldsymbol{\theta}$.

The Bayesian optimal design, $\mathbf{d}^*$, maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$ with respect to the future data $\mathbf{y}$ and model parameter $\boldsymbol{\theta}$:

$$\begin{aligned} \mathbf{d}^* &= \arg\max_{\mathbf{d}\in\mathbf{D}} \; E\{U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})\} \\ &= \arg\max_{\mathbf{d}\in\mathbf{D}} \; \int_{\mathbf{Y}} \int_{\boldsymbol{\Theta}} U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}) d\boldsymbol{\theta} d\mathbf{y}. \end{aligned} \tag{5.1}$$

The integration is performed over the sample space $\mathbf{Y}$ of the data, and the parameter space $\boldsymbol{\Theta}$.

Unless the likelihood and prior are specifically chosen to enable analytic evaluation of the integration problem, equation (5.1) does not usually have a closed form solution. Therefore, numerical approximations or stochastic solution methods are required to solve the maximisation and integration problem.

Bayesian optimal design methods should be employed when the experimenter wishes to perform a Bayesian analysis on the data that is collected using the experimental design.

5.1.2 Pharmacokinetic Studies

Pharmacokinetic (PK) studies investigate the disposition of a drug following its administration to a subject or group of study subjects. PK studies generally assume that the change in drug concentration over time can be described by a particular model, such as a compartmental model. Compartmental models are usually derived by solving a series of ordinary differential equations and error terms are incorporated into the model to account for any systematic or natural variation that may be present in the data. PK studies are often interested in measures such as the area under the concentration-time curve (AUC), maximum concentration (C_{max}), time of maximum concentration (t_{max}), elimination half-life ($t_{1/2}$), clearance rate (CL), and volume of distribution (V) (e.g., Alderman et al. (1998); Stroud et al. (2001); Hiemenz et al. (2005); Saint-Marcoux et al. (2005)).

During PK studies, one cannot directly observe the kinetics of the drug in the study subjects, and so samples are instead taken from biological fluids such as blood, plasma or urine. These samples are taken at specific times and the drug and metabolite concentrations are measured. The choice of plasma sampling times is highly important in PK studies. One should avoid complex designs that require a large number of samples to be taken from each study subject, since this would be costly and inconvenient for the study

subjects. Thus, PK studies require the timing and number of samples to be carefully planned, so as to gain accurate estimates of the parameters but also prevent physical and mental strain on the study subjects, and reduce study costs.

Atkinson et al. (1993) found designs which minimised the variance of the AUC, C_{max}, and t_{max} estimates for an open one-compartmental PK model with first-order absorption input and a constant variance term. Prior distributions were used to account for parameter uncertainty. The designs found by Atkinson et al. (1993) were c_θ-optimum and D_θ-optimum designs, and were 'pseudo' Bayesian designs, since the utility functions were based on the Fisher information matrix. We will extend the work of Atkinson et al. (1993) to enable fully Bayesian designs to be found for several of these PK parameters of interest.

5.1.3 Extracorporeal Membrane Oxygenation

Extracorporeal membrane oxygenation (ECMO) is a prolonged form of cardiopulmonary bypass and involves the diversion of blood flow through a circuit located outside of the body so that the blood may be oxygenated. ECMO is used as a final resort in critically ill patients with potentially reversible respiratory failure to temporarily support the heart and/or lungs. In 2009, during the worldwide H1N1 pandemic, ECMO was a vital treatment modality for H1N1 patients requiring advanced ventilatory support (Davies et al. (2009); Hui et al. (2010); Napolitano et al. (2009); Chang et al. (2010)), and has a reported survival rate of approximately 50% when used in critically ill adults with respiratory failure (Brogan et al. (2009)). Despite the major benefits that ECMO can afford, it can also induce a wide variety of PK changes in critically ill patients that are not fully understood, and can lead to therapeutic failure or toxicity. Thus there is a need to understand the PK of critically ill patients undergoing ECMO.

5.1.4 Contribution and Outline

Optimal Bayesian experimental design involves sampling from the posterior distribution for many possible future data sets that are drawn from the prior predictive distribution. Therefore, many thousands of posterior distributions are required, and fast methods for approximating the posterior are necessary so that computation can be performed in a reasonable amount of time. Importance sampling has commonly been used in the literature to rapidly obtain samples from the posterior in the context of Bayesian experimental design (Kinas (1996); Cook et al. (2008); Dror and Steinberg (2008); McGree et al. (2012c); Ryan et al. (2014c)), where the prior is used as the importance distribution. However, importance sampling from the prior tends to break down and is inefficient if there is a reasonable number of experimental observations since the posterior distribution can be very different from the prior. Importance sampling is also inefficient if the model parameter is high dimensional. In this paper, we explore the use of the Laplace approximation for calculating Bayesian utility functions to overcome this drawback. Furthermore, we consider using the Laplace approximation to form the importance distribution to obtain a more efficient importance distribution than the prior.

The methodology is motivated by a PK study conducted by Shekar et al. (2013) which investigates the effect of ECMO on the PK of antibiotics in sheep. Here we will re-design their study to find 10 near optimal plasma sampling times which produce precise estimates of pharmacokinetic model parameters/measures of interest. In the PK study (Shekar et al. (2013)), healthy sheep had been administered the antibiotic, meropenem, and also underwent ECMO treatment. PK measurements were taken both before the sheep had undergone ECMO, and after ECMO treatment had commenced, to determine the effect of ECMO on certain PK parameters of interest. We consider different utility functions in these studies, which involve posterior distributions of these PK parameters.

Whilst the algorithms we have borrowed from the Bayesian inference literature are not novel, to our knowledge, no previous works have investigated and compared the use of importance sampling and Laplace approximations for estimating the posterior in a Bayesian experimental design context. These estimates of the posterior are used to calculate Bayesian utility functions, some of which were specifically developed for our design problem of interest. Such a methodological investigation is necessary in order to advance the experimental design literature so that optimal and fully Bayesian designs can be found when there is a large number of sampling times. The importance of this methodological development is highlighted by the ECMO application considered in this paper.

In Section 5.2 we introduce the motivating case study. Section 5.3 describes the utility functions used in this paper and the methods we use for estimating them. Our design methodology is outlined in Section 5.4 and our methods are applied to the case study in Section 5.5. The article concludes with a discussion in Section 5.6.

5.2 Case Study: Determining Sampling Times for a Study Investigating the Effects of ECMO on the PK of Meropenem in Sheep

Our design problem is based on a PK study conducted by Shekar et al. (2013), in which healthy sheep were used as their own controls. Baseline PK data for meropenem (and other study drugs) were obtained from two healthy sheep, prior to commencing (venovenous) ECMO. Once ECMO began, 500 mg of meropenem (and the other study drugs) was infused over 30 minutes and blood samples were taken at 0, 15, 30, 45, 60, 90, 120, 180, 360, 480 and 720 minutes after the commencement of the drug infusion. These blood samples were taken at the same time as the baseline PK measurements (when the sheep were not on ECMO). Here we will be conducting a 'retrospective study design' using existing PK data, to determine whether the study design can be improved upon for future studies.

It should be noted that our experimental design goal is to determine the 10 (near) optimal blood sampling times that give rise to precise estimates of PK parameters of interest. For reasons that will be clarified in Section 5.3, we are interested in using a somewhat large number of sampling times, and are not concerned with reducing the number of observations taken. If one were interested in reducing the resources involved in the study,

due to time or cost constraints, then one could investigate the optimal number of design points (e.g., Sheiner and Beal (1983); Dror and Steinberg (2008)).

The data is assumed to be modelled by a one-compartment infusion PK model with fixed effects. This model does not account for individual variability since our motivating case study only had data (which was very similar) available for two sheep. The model consists of two parameters: the volume of distribution V, which is a theoretical volume that a drug would have to occupy to provide the same concentration as is currently present in the blood plasma (if the drug were uniformly distributed), and the first-order elimination rate constant k_e. If y_t denotes the observed concentration at time t minutes following the administration of the drug, then the model may be given by:

$$y_t = \begin{cases} \frac{D}{T_{inf}} \frac{1}{k_e V}(1 - e^{-k_e t}) \cdot (1 + \epsilon_t) & \text{if } t \leq T_{inf} \\ \frac{D}{T_{inf}} \frac{1}{k_e V}(1 - e^{-k_e T_{inf}}) e^{-k_e (t - T_{inf})} \cdot (1 + \epsilon_t) & \text{if } t > T_{inf}, \end{cases}$$

where $\epsilon_t \sim N(0, \sigma^2_{\text{prop}})$, and $t \in [0, 720]$ minutes.

Here the dose, D, is 500 mg, which was administered over a period of 30 minutes ($T_{inf} = 30$). Only proportional error ϵ_t (and not additive error) is present in the model. That is, the variance is not constant and depends on the mean concentration at time t. This error structure gave a better fit to the data compared to models that contained only additive error, both additive and proportional error, or exponential error. A Metropolis-Hastings Markov chain Monte Carlo (MH MCMC) algorithm was used to fit the above model to existing data for 1 of the sheep that underwent ECMO. The MCMC samples of the PK parameters that resulted from fitting the data to the model (Figure 5.1 (a) and 5.1 (b)) were used as the prior distributions for the retrospective design:

$$\boldsymbol{\theta} \sim \text{MVN}\left\{ \begin{pmatrix} -3.26 \\ 8.99 \end{pmatrix}, \begin{pmatrix} 0.0071 & -0.0057 \\ -0.0057 & 0.0080 \end{pmatrix} \right\},$$

where $\boldsymbol{\theta} = (\log k_e, \log V)$. Based on the MCMC samples (Figure 5.1 (c)), σ^2_{prop} was assumed to follow a gamma distribution with mean 0.01 and variance 10^{-5}. σ^2_{prop} was considered a nuisance parameter that was independent of k_e and V, and was not of interest to estimate. The prior predictive curves of the data are displayed in Figure 5.2.

Here our design points $\mathbf{d}$ are the sampling times $\mathbf{t}$ (in minutes), and we are searching for 10 (near) optimal sampling times in addition to $t = 0$.

5.3 Bayesian Utility Functions and Their Estimation

In this article we investigate various design criteria that are concerned with the precise estimation of PK parameters. These design criteria are Bayesian design criteria and assume that a Bayesian analysis will be performed on any data that is generated from the experimental design. Bayesian utility functions are typically a function of $\mathbf{d}$ and $\mathbf{y}$, but not $\boldsymbol{\theta}$. Equation (5.1) does not typically have a closed form. Therefore we will use Monte Carlo methods and/or Laplace approximations (described in Section 5.3.2)

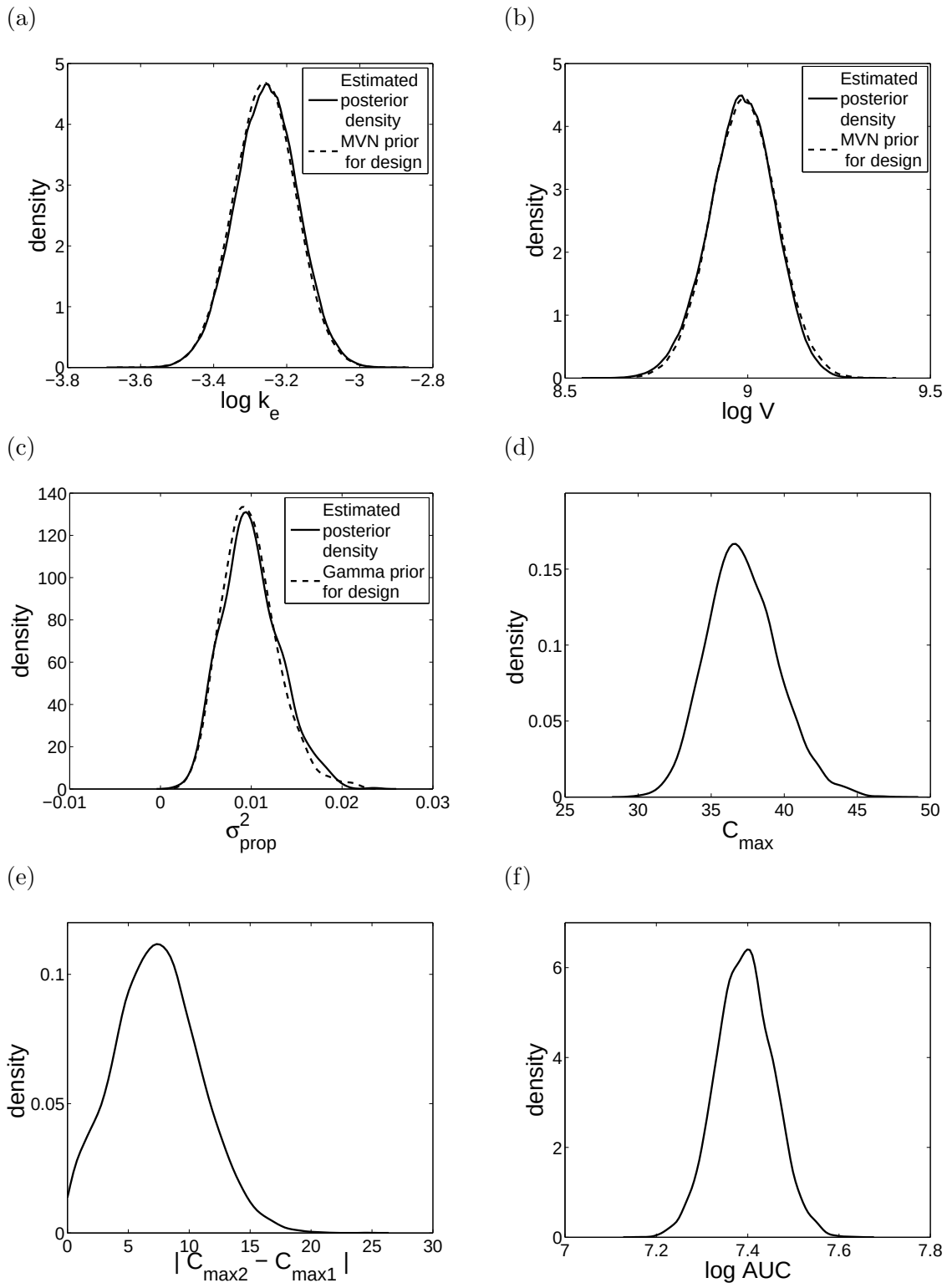


Figure 5.1: Estimated posterior distributions of PK parameters of interest that resulted from the one-compartment PK model being fitted to the data via MCMC simulations: (a) estimated posterior density of $\log k_e$ (continuous line) with the prior for design overlaid (dotted line), (b) estimated posterior density of $\log V$ (continuous line) with the prior for design overlaid (dotted line), (c) estimated posterior density of σ^2_{prop} (continuous line) with the prior for design overlaid (dotted line), (d) estimated posterior density of C_{max}, (e) estimated posterior density of the absolute difference in C_{max} between sheep on ECMO and not on ECMO, and (f) estimated posterior density of the log AUC. These posterior samples were used as priors for our design problem of interest.

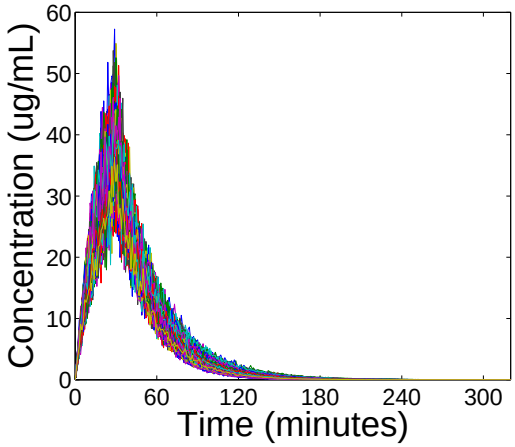


Figure 5.2: Prior predictive curves of the 1 compartment infusion PK model (online version in colour).

to approximate the posterior distribution. Once an approximation of the posterior has been obtained, it is typically straightforward to estimate the utility functions. For all of the utility functions mentioned in Section 5.3.1, we are interested in finding the optimal design $\mathbf{d}^*$, that maximises the expected utility function $U(\mathbf{d})$ over the design space $\mathbf{D}$, with respect to the unknown data $\mathbf{y}$ and model parameter $\boldsymbol{\theta}$.

5.3.1 Utility functions

Posterior Precision of the Peak Concentration Estimate

Pharmacologists are often interested in determining the maximum concentration of a drug (C_{max}) in patients during PK studies to ensure that patients are not exposed to toxic concentrations (e.g., Alderman et al. (1998); Stroud et al. (2001); Saint-Marcoux et al. (2005)). The posterior precision of the peak concentration estimate may be expressed as follows:

$$U_1(\mathbf{d}, \mathbf{y}) = \{\text{Var}(\phi_1|\mathbf{d}, \mathbf{y})\}^{-1},$$

where ϕ_1 represents C_{max}. C_{max} occurs at the time when the infusion of the drug is complete (i.e., when $t = T_{inf}$). The expression for C_{max} may be given by:

$$\phi_1 = C_{max} = y(T_{inf}) = \frac{D}{T_{inf}} \frac{1}{k_e V}(1 - e^{-k_e T_{inf}}). \tag{5.2}$$

For our case study, we decided to look at the C_{max} of meropenem for sheep receiving ECMO treatment. The posterior distribution of C_{max} that resulted from the MCMC fit to the data is displayed in Figure 5.1 (d).

Posterior Precision of the Difference in Peak Concentration Estimates Between Sheep on ECMO and not on ECMO

Since pharmacologists are interested in determining the effect of ECMO on the PK curve (Shekar et al. (2013)), we decided to determine the (near) optimal design points which maximise the precision of the estimate of the difference in the concentration-time curve

peaks for sheep on ECMO versus sheep not receiving ECMO treatment. The posterior precision of the difference in the peak concentration estimates between sheep on ECMO and sheep not receiving ECMO treatment may be expressed as:

$$U_2(\mathbf{d}, \mathbf{y}) = \{\text{Var}(\phi_2|\mathbf{d}, \mathbf{y})\}^{-1},$$

where ϕ_2 represents the absolute difference in the posterior peak concentrations for sheep not on ECMO and sheep on ECMO. The peak concentrations were calculated using equation (5.2), and then the absolute difference in the peaks was found:

$$\phi_2 = |C_{max_2} - C_{max_1}| = \left| \frac{D}{T_{inf}} \frac{1}{k_{e_2} V_2} (1 - e^{-k_{e_2} T_{inf}}) - \frac{D}{T_{inf}} \frac{1}{k_{e_1} V_1} (1 - e^{-k_{e_1} T_{inf}}) \right|,$$

where $\boldsymbol{\theta}_1 = (\log k_{e_1}, \log V_1)$ are the PK parameters for when the (one) sheep was on ECMO and $\boldsymbol{\theta}_2 = (\log k_{e_2}, \log V_2)$ are the PK parameters for when the (one) sheep was not on ECMO. The prior for $\boldsymbol{\theta}_1$ is given in Section 5.2. The prior for $\boldsymbol{\theta}_2$ is:

$$\boldsymbol{\theta}_2 \sim \text{MVN} \left\{ \begin{pmatrix} -3.59 \\ 8.94 \end{pmatrix}, \begin{pmatrix} 0.0055 & -0.0045 \\ -0.0045 & 0.0062 \end{pmatrix} \right\}.$$

The parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are assumed to be independent a priori. The posterior distribution of the absolute difference in C_{max} (between sheep on ECMO and sheep not on ECMO) that resulted from the MCMC fit to the data is displayed in Figure 5.1 (e).

Posterior Precision of the (log) AUC Estimate

Pharmacologists are also interested in estimating the AUC for concentration-time curves to determine the patients' total exposure to a drug (e.g., Alderman et al. (1998); Stroud et al. (2001); Hiemenz et al. (2005); Saint-Marcoux et al. (2005)). To enable an analytic solution to be found for the precision of the posterior distribution of this parameter of interest (when using Laplace approximations to approximate the posterior distribution), we decided to instead look at the log AUC, which was closer to being normally distributed a priori than the AUC (non-transformed).

Again, we can express the utility function as:

$$U_3(\mathbf{d}, \mathbf{y}) = \{\text{Var}(\phi_3|\mathbf{d}, \mathbf{y})\}^{-1},$$

where ϕ_3 represents the log AUC. Since we have a (simple) one-compartment PK model, we can find the log AUC analytically, by $\log \text{AUC} = \log D - \log k_e - \log V$. Therefore,

$$\phi_3 = \log D - \log k_e - \log V.$$

For our case study, we decided to look at the log AUC of meropenem for sheep receiving ECMO treatment. The posterior distribution of the log AUC that resulted from the MCMC fit to the data is displayed in Figure 5.1 (f).

Determinant of the Posterior Variance-covariance Matrix

If one is interested in designing for the precise estimation of the elimination constant $\log k_e$ and the volume of distribution $\log V$, then one could use the inverse of the determinant of the posterior variance-covariance matrix of the PK model parameters as the utility function. The inverse of the determinant of the posterior variance-covariance matrix of the model parameter is also known as the 'Bayesian D-posterior precision' (Drovandi et al. (2013)) and is given by:

$$U_4(\mathbf{d}, \mathbf{y}) = \frac{1}{\det(\text{Var}(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y}))}.$$

This utility is estimated by finding the reciprocal of the determinant of the variance-covariance matrix of the $\boldsymbol{\theta}$ sample from the posterior. For our case study, we looked at precisely estimating $\log k_e$ and $\log V$ for sheep receiving meropenem and on ECMO treatment.

5.3.2 Methods for Estimating Utility Functions

Each of the above-mentioned utility functions requires the posterior distribution $p(\boldsymbol{\theta}|\mathbf{d}, \mathbf{y})$. However, the posterior often does not have a closed form expression and so numerical methods are required to sample from or approximate the posterior distribution. Each possible future data set that is drawn from the prior predictive distribution requires calculations of the posterior distribution, and so many thousands of posterior distributions need to be considered. Hence, fast methods for obtaining the posterior distribution are required. In this paper we explore, compare and contrast importance sampling and Laplace approximations for this purpose.

Importance Sampling

Importance sampling is a commonly-used approach for approximating target distributions of interest (Geweke (1989)). It involves choosing an alternative distribution $g(\cdot)$ (the *importance* distribution), from which it is easy to sample, then appropriately weighting the samples to account for the discrepancy between $g(\cdot)$ and the target distribution. Here the target distribution is the posterior $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$. This produces weighted samples $\{\boldsymbol{\theta}_k, W_k\}_{j=1}^{M_p}$, where M_p is the number of particles used to approximate $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$, $w(\boldsymbol{\theta}) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{d})p(\boldsymbol{\theta})}{g(\boldsymbol{\theta})}$ are the importance weights, and $W_k \propto w(\boldsymbol{\theta}_k)$ are the normalised importance weights, $\sum_{k=1}^{M_p} W_k = 1$. The distributions $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{d})$ and $g(\cdot)$ should have the same support. A common approach in Bayesian experimental design is to use the prior as the importance distribution $g(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ (e.g., Cook et al. (2008); Dror and Steinberg (2008); Ryan et al. (2014c)), and the importance weights are reduced to the likelihood function. However, this can be very inefficient for diffuse priors (Chopin (2002)) or concentrated likelihoods, which result here from having several observations in the design.

To measure the efficiency of importance sampling, the effective sample size (ESS) is used, where

$$ESS = \frac{1}{\sum W_k^2}, 1 \leq ESS \leq M_p.$$

For this work, we have used the following importance sampling scheme to approximate the posterior distribution. Let the utility function, which is the posterior precision of some parameter of interest, $\phi(\boldsymbol{\theta})$, be denoted by:

$$U(\mathbf{d}, \mathbf{y}) = \{\text{Var}(\phi(\boldsymbol{\theta})|\mathbf{d}, \mathbf{y})\}^{-1}.$$

The posterior variance can be computed using the following estimator (see Stroud et al. (2001)):

$$\widehat{\text{Var}}(\phi(\boldsymbol{\theta})|\mathbf{d}, \mathbf{y}) = \sum_{k=1}^{M_p} (\phi(\boldsymbol{\theta}_k) - \bar{\phi})^2 W_k,$$

where $\{\boldsymbol{\theta}_k\}_{k=1}^{M_p} \sim p(\boldsymbol{\theta})$ are samples drawn from the prior (i.e., the prior is used as the importance distribution), and the importance weights are proportional to the likelihood of the observed data,

$$w_k \propto p(\mathbf{y}|\boldsymbol{\theta}_k, \mathbf{d}), \quad \text{for } k = 1, ..., M_p,$$

and the estimate of the posterior mean, $\bar{\phi}$, is given by

$$\bar{\phi} = \sum_{k=1}^{M_p} \phi(\boldsymbol{\theta}_k) W_k.$$

This approach was used to calculate the following utility functions: the posterior precision of C_{max}; the posterior precision of the difference in C_{max} between sheep on ECMO and sheep not on ECMO; and the posterior precision of the log AUC estimate.

To estimate the inverse of the determinant of the posterior variance-covariance matrix of the PK parameters, variances and covariances were estimated based on this weighted sample and used to form the variance-covariance matrix. For each of these utility functions, the proportional variance σ^2_{prop} was only used to calculate the importance weights, and was not one of the parameters of interest that we were designing to precisely estimate.

In our applications, the number of particles M_p was chosen to ensure that reasonably stable (based on the ESS) and precise estimates of the utility were obtained. For each iteration of the MCMC algorithm that was used (see Section 5.4.2), the M_p value was increased until an ESS value of 1000 or more was obtained, so that the utilities could be estimated using at least the equivalent of 1000 independent samples from the posterior. We conducted a sensitivity analysis into this minimum value of the ESS, and found that values less than 1000 tended to give less precise estimates of the utility and did not provide a good exploration of the design space as the utility function would become 'stuck' at a particular value during the search over the design space.

When the inverse of the determinant of the posterior variance-covariance matrix was used as the utility function, a large number of samples was often required to ensure an ESS value of 1000. However, an upper bound of $M_p = 100$ million had to be set, as larger values than this caused memory storage issues. Therefore, some of the samples which were used to calculate this utility function had an ESS value less than 1000. This highlights the

difficulty of using the prior as the importance distribution for designs with a reasonable number of observations.

Importance sampling from the prior is usually inefficient for large amounts of data since the posterior distribution is very different from the prior. Although reducing the number of observations taken would improve the performance of importance sampling, this lies outside the scope of the current study, as we are not concerned with time or cost constraints. In this study, we are interested in comparing the performance of importance sampling to other methods for approximating the posterior distribution (see below) for design problems where (somewhat) large amounts of data are involved.

Laplace Approximation

It is widely known that Monte Carlo methods, such as importance sampling, may fail in cases where large amounts of data are involved and/or the model parameter is high dimensional (e.g., Bengtsson et al. (2008); Ryan et al. (2014c)). To overcome this, we suggest the use of Laplace approximations to the posterior distribution of $\boldsymbol{\theta}$ (suitably parameterised), instead of importance sampling. Here the Laplace approximation assumes a multivariate normal distribution for the posterior distribution of the parameter $\boldsymbol{\theta}$, as well as $\log \sigma^2_{\text{prop}}$. One simply uses a numerical search algorithm, such as a quasi-Newton method, to estimate the posterior mode

$$(\hat{\boldsymbol{\theta}}, \log \hat{\sigma}^2_{\text{prop}}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta},\, \log\sigma^2_{\text{prop}}\in\text{Re}} \log p(\boldsymbol{\theta}, \log \sigma^2_{\text{prop}}|\mathbf{y}, \mathbf{d}),$$

and the posterior variance-covariance matrix (of $\boldsymbol{\theta}$) via the upper-left quadrant of -1 times the inverse of an estimate of the Hessian matrix evaluated at the $(\hat{\boldsymbol{\theta}}, \log \hat{\sigma}^2_{\text{prop}})$ value. The approximation can be used directly (as we do for the Bayesian D-posterior precision utility) or samples can be drawn from this Gaussian approximation to estimate quantities that are functions of the model parameters (e.g., C_{max}). However, the estimated posterior distribution obtained from the Laplace approximation may not accommodate the tails of the posterior distribution, and so we investigate alternative methods to obtain better coverage of the tails.

Importance Sampling with a More Informed Importance Distribution ("Combined Approach")

To improve the efficiency of the prior as the importance distribution (which may be too diffuse), we considered the use of the Laplace approximation to form the importance distribution. To obtain better coverage of the tails, we multiplied the variance-covariance matrix that was obtained from the Laplace approximation by 2 and drew many samples from the multivariate normal distribution, calculated quantities such as the AUC or C_{max}, and then weighted these samples by the importance weights. This approach may also account for any non-normality in the posterior distribution. The importance weights were calculated in a similar manner to importance sampling from the prior but with one major difference: previously, the importance distribution $g(\boldsymbol{\theta})$ was the prior distribution,

but for the combined approach, the importance distribution is the multivariate normal density which resulted from the Laplace approximations.

Other Methods

Previous approaches, such as Han and Chaloner (2004), have used MCMC to approximate the posterior distribution for the utility function calculations. However, Han and Chaloner (2004) only investigated fixed designs and no optimisation was performed over the design space. Whilst MCMC is useful and often appropriate for Bayesian data analysis, it may not be suitable for optimal Bayesian experimental design as it is computationally intensive to perform MCMC to approximate the posterior distribution for each of the thousands of iterations required in the Bayesian experimental design algorithms. For importance sampling, the precision of the algorithm can be controlled by using the ESS. An ESS is more difficult to determine for MCMC; a burn-in period is required and so it is more difficult to automate; and various tests are required to determine convergence of the MCMC algorithm. For these reasons we did not investigate the use of MCMC to approximate the posterior distributions required to perform Bayesian experimental design.

We note that a multivariate t-distribution could be used in place of the multivariate normal distribution for the importance distribution in the "combined approach" to obtain a wider coverage of the tails. However, we found our approach (where the variance-covariance matrix of the multivariate normal distribution is twice the value of the variance-covariance matrix obtained from the Laplace approximation) to be sufficient for our motivating application. It is important to note that use of the Laplace approximation relies on the assumption that the posterior distribution of the parameter $\boldsymbol{\theta}$ is reasonably well approximated by a multivariate normal distribution. If this assumption is not valid, then use of Laplace approximations to estimate the posterior distribution may not be appropriate.

5.4 Design Methodology

5.4.1 Stochastic Optimisation

One of the most commonly-used stochastic algorithms for solving the maximisation and integration problem (equation (5.1)) involves the use of Monte Carlo simulations to evaluate the integral(s) (equation (5.1)) (e.g., Müller (1999); Huan and Marzouk (2012)). In the majority of situations, $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d})$ is available for efficient random variable generation and the utility function can be evaluated point-wise using the simulated $(\boldsymbol{\theta}_i, \mathbf{y}_i)$ for $i = 1, ..., M$. The integral may then be approximated by using:

$$\hat{U}(\mathbf{d}) = \frac{1}{M} \sum_{i=1}^{M} U(\mathbf{d}, \boldsymbol{\theta}_i, \mathbf{y}_i). \tag{5.3}$$

One can then use $\hat{U}(\mathbf{d})$ to find the optimal design, $\mathbf{d}^* = \arg\max \hat{U}(\mathbf{d})$, by using a suitable maximisation method (see Müller (1999)). However, straightforward Monte Carlo

integration over $(\boldsymbol{\theta}, \mathbf{y})$ for each design $\mathbf{d}$ can be computationally intensive for high dimensional design problems since a large value of M is required to obtain reasonable accuracy of the estimate of $U(\mathbf{d})$.

Instead, Müller and colleagues (Clyde et al. (1996); Bielza et al. (1999); Müller (1999)) performed the integration using MCMC which sampled from the target distribution:

$$h(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) \propto U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}),$$

using a MH MCMC scheme. It was assumed that $U(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$ satisfies the appropriate conditions for $h(\cdot)$ to be positive and integrable over $(\mathbf{D}, \boldsymbol{\Theta}, \mathbf{Y})$. The probability distribution $h(\cdot)$ is defined such that the marginal distribution of $\mathbf{d}$ is proportional to the expected utility, $U(\mathbf{d})$. The MCMC simulation focuses on sampling in areas of high expected utility and discourages sampling in areas of low expected utility (see Müller (1999)). The sample of simulated $\mathbf{d}$ may be used to provide an estimate of $h(\mathbf{d})$ and the joint mode of $h(\mathbf{d})$, $\mathbf{d}^*$, corresponds to the optimal design.

A 'simulated annealing-type approach' (see Van Laarhoven and Aarts (1987); Clyde et al. (1996); Bielza et al. (1999); Müller (1999)) has also been suggested, in which a sample is simulated from:

$$h_J(\mathbf{d}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_J, \mathbf{y}_1, ..., \mathbf{y}_J) \propto \prod_{j=1}^{J} U(\mathbf{d}, \boldsymbol{\theta}_j, \mathbf{y}_j) p(\boldsymbol{\theta}_j, \mathbf{y}_j|\mathbf{d}).$$

The marginal $h_J(\mathbf{d})$ is proportional to $U^J(\mathbf{d})$, where J is an integer. As J increases, the utility surface will become more peaked and simulations will cluster more tightly around the mode.

Alternative simulation-based algorithms for maximising the expected utility, such as sequential Monte Carlo (see Amzal et al. (2006)), have also been proposed. However, we found the MCMC approach of Müller and colleagues (Clyde et al. (1996); Bielza et al. (1999); Müller (1999)) sufficient for sampling the design space for our design problems.

5.4.2 MCMC Algorithm

To solve the optimal design problems (equation (5.1)), we implemented the MH MCMC algorithm presented by Müller (1999), which is described in Algorithm 5.1 in Appendix A, to perform simulation from $h(\cdot)$. We found that $J = 5$ provided sufficiently peaked utility surfaces for our design problems.

We note that the joint mode of $h(\mathbf{d})$ needs to be found rather than the marginal modes for each element of $\mathbf{d}$ as the latter may be very different from the former.

MCMC runs of 10 000 iterations were performed and the convergence of the algorithms was carefully monitored (through examination of autocorrelation plots, histograms and contour plots of the design variables, and trace plots of the utility functions over the iterations). To determine the optimal designs, we searched for the multivariate mode of

the multivariate normal kernel smoothing density estimates of the design variables (see Cook et al. (2008); Drovandi and Pettitt (2013)).

5.4.3 Lower Dimensional Parameterisations

Here we use an approach from our previous work (Ryan et al. (2014c)) that reduces the computational burden of searching for a large number of design points. This approach involves the use of lower dimensional parameterisations that consist of a few design variables, which generate multiple design points. Using this approach, one simply has to search over a few design variables, rather than searching for a large number of optimal design points, thus providing substantial computational savings. The following lower dimensional parameterisation schemes were used to generate the design points:

(a) $d_s = d_1\delta^{(s-1)}$, where $d_1 \geq 0, \delta > 1$ ('geometric scheme');
(b) $d_s = d_1 + \delta \times (s-1)$, where $d_1 \geq 0, \delta > 0$ ('even spacing scheme'); and
(c) Percentiles of a Beta(a, b) distribution, scaled to [0,T] - the design space, where $a, b > 0$ ('beta scheme').

Here d_1 is the first design point, δ is a spacing parameter, (a, b) are positive shape parameters for a beta distribution, s is an index where $s = 1, ..., n_d$ and n_d is the number of design points. Under these lower dimensional parameterisations of the design points, the Müller (1999) algorithm searched over two design variables (d_1, δ), or (a, b), depending on the scheme that was used. This avoids the need to search for a large number of optimal design points and provides substantial computational savings. Also, it is much easier to obtain the multivariate mode for a few design variables than for a large number of design variables. However, it should be stressed that the designs generated by the lower dimensional parameterisations are not optimal but *near* optimal, which is a compromise of the computational savings achieved through these methods.

A typical experimental design for PK studies where the drug concentration-time profile is thought to be modelled by a compartmental model is the geometric design ('geometric scheme' above) (Atkinson et al. (1993)). We also decided to investigate the use of evenly-spaced designs and designs which arose from the (evenly-spaced) percentiles of the beta distribution as we thought that these schemes may give fairly flexible designs that could be suitable for use in PK studies.

For comparative purposes, we also searched for optimal designs for a three design (support) point problem using the Müller (1999) algorithm, where the design variables were (d_1, d_2, d_3), i.e., the three sampling times for the PK study. This involved searching for three optimal sampling times which were generated in the MH MCMC algorithm via normal random walks, and did not involve the use of the lower dimensional parameterisation schemes. Once these three sampling times were determined, replicates were placed on each of these support points so that a total of 10 design points were obtained (4 replicates on d_1, 3 replicates on d_2, 3 replicates on d_3). Since true replication is not practically feasible for a PK study, these 'replicates' were separated by a time interval of 15 minutes.

Following the dimension reduction of the design problem (to two design variables), a brute force approach was also investigated, in which the expected utility was calculated using Monte Carlo integration (equation (5.3)) and optimisation was performed by searching over a grid of values for the design variables (e.g., Müller and Parmigiani (1995)). However, the brute force approach was found to be too computationally intensive and it was difficult to determine an appropriate grid of values to encompass regions of high expected utility for our applications of interest. Therefore, results will not be discussed for this method.

All simulations were performed on an Intel Xeon Processor E5-2670 (2.66 GHz, 1 GB RAM).

5.5 Results

The utility function values for the optimal designs (Figures 5.3 - 5.6) were calculated using Monte Carlo integration (equation (5.3)) with $M = 1000$. The Monte Carlo sample size was chosen to ensure that the 95% CIs were accurate to two decimal places. Figures 5.3 - 5.6 appear in colour in the online version of this article.

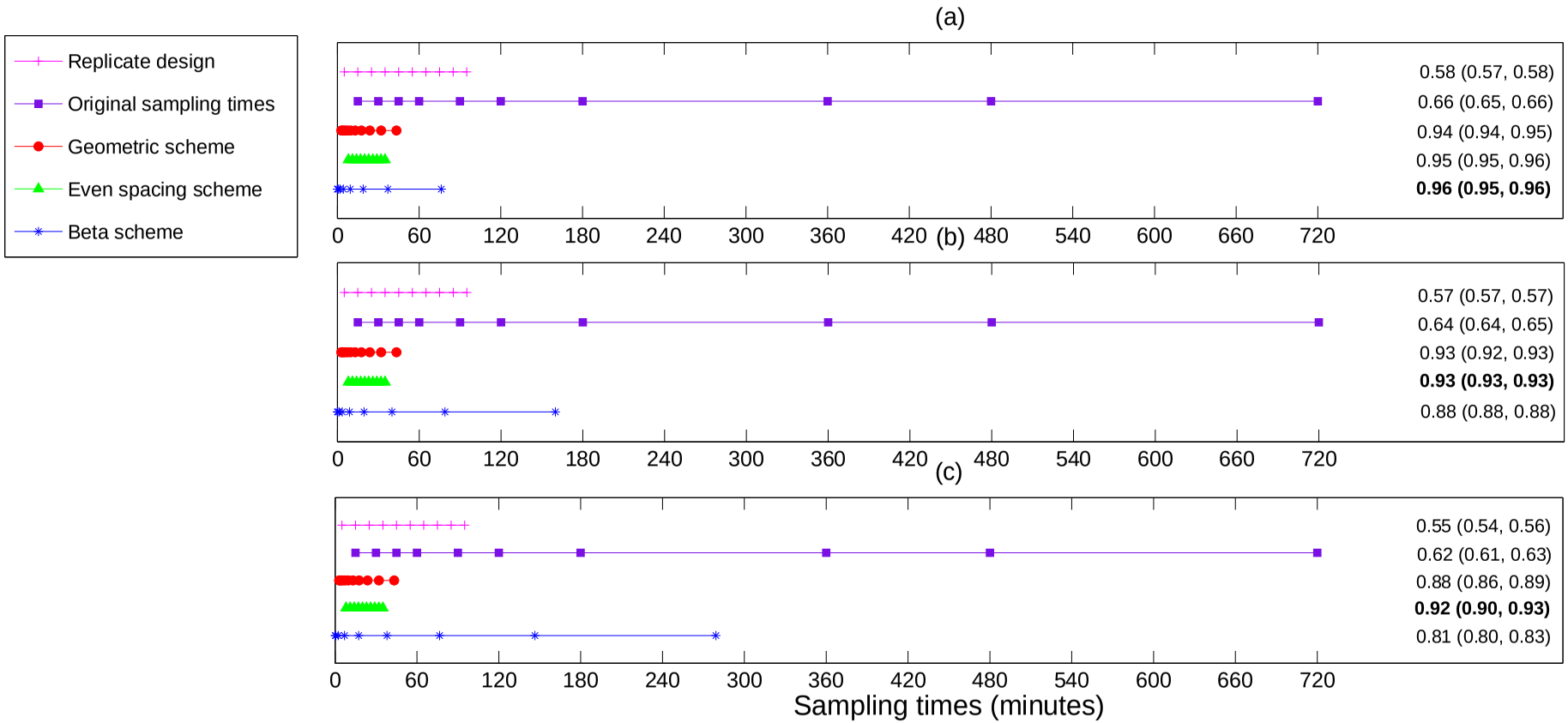


Figure 5.3: The utility is the posterior precision of C_{max}: PK sampling times generated under the three lower dimensional parameterisation schemes, as well as the replicate designs and original designs, found using (a) importance sampling, (b) Laplace approximations, and (c) combination of Laplace approximations and importance sampling to estimate the posterior precision of C_{max} utility function.

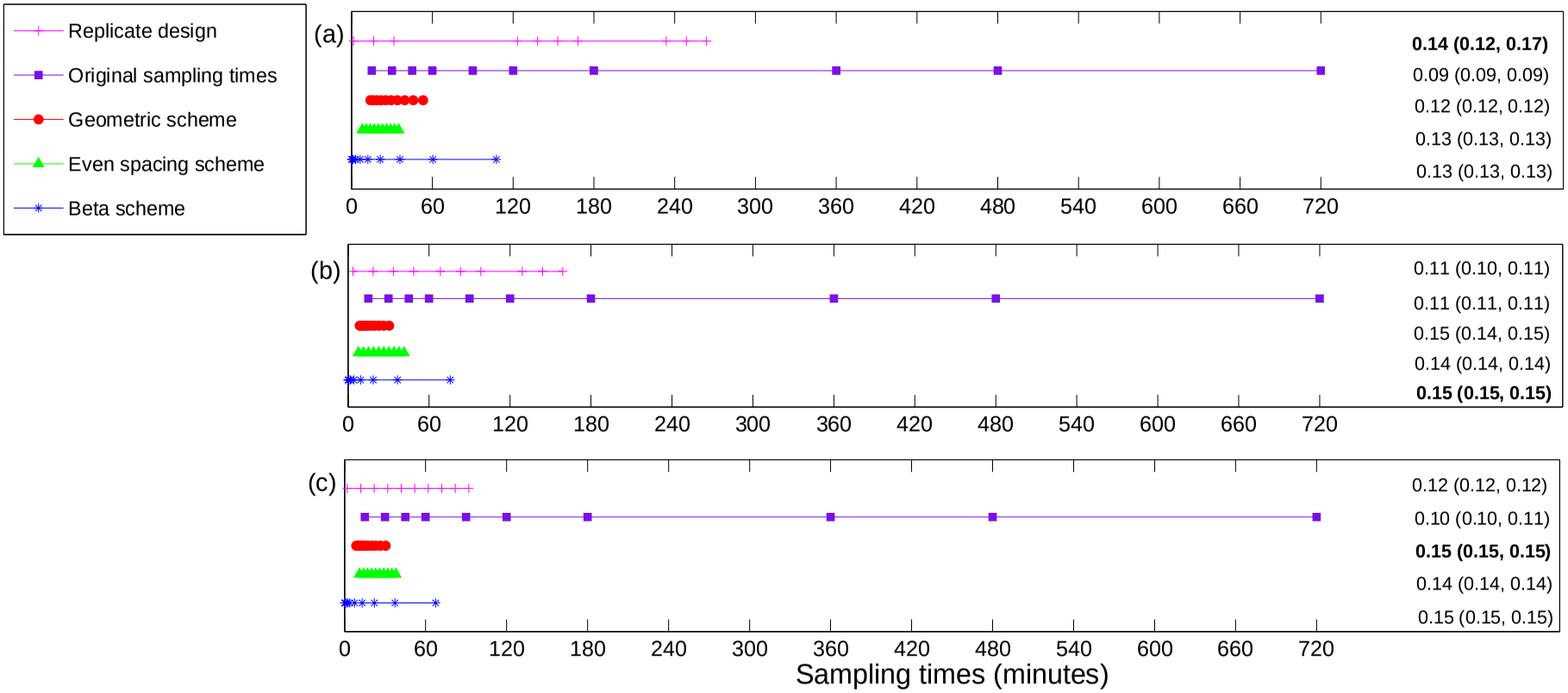


Figure 5.4: The utility is the posterior precision of absolute difference in C_{max} for sheep on ECMO vs sheep not on ECMO: PK sampling times generated under the three lower dimensional parameterisation schemes, as well as the replicate designs and original designs, found using (a) importance sampling, (b) Laplace approximations, and (c) combination of Laplace approximations and importance sampling to estimate the posterior precision of absolute difference in C_{max} estimates for sheep on ECMO vs sheep not on ECMO utility function.

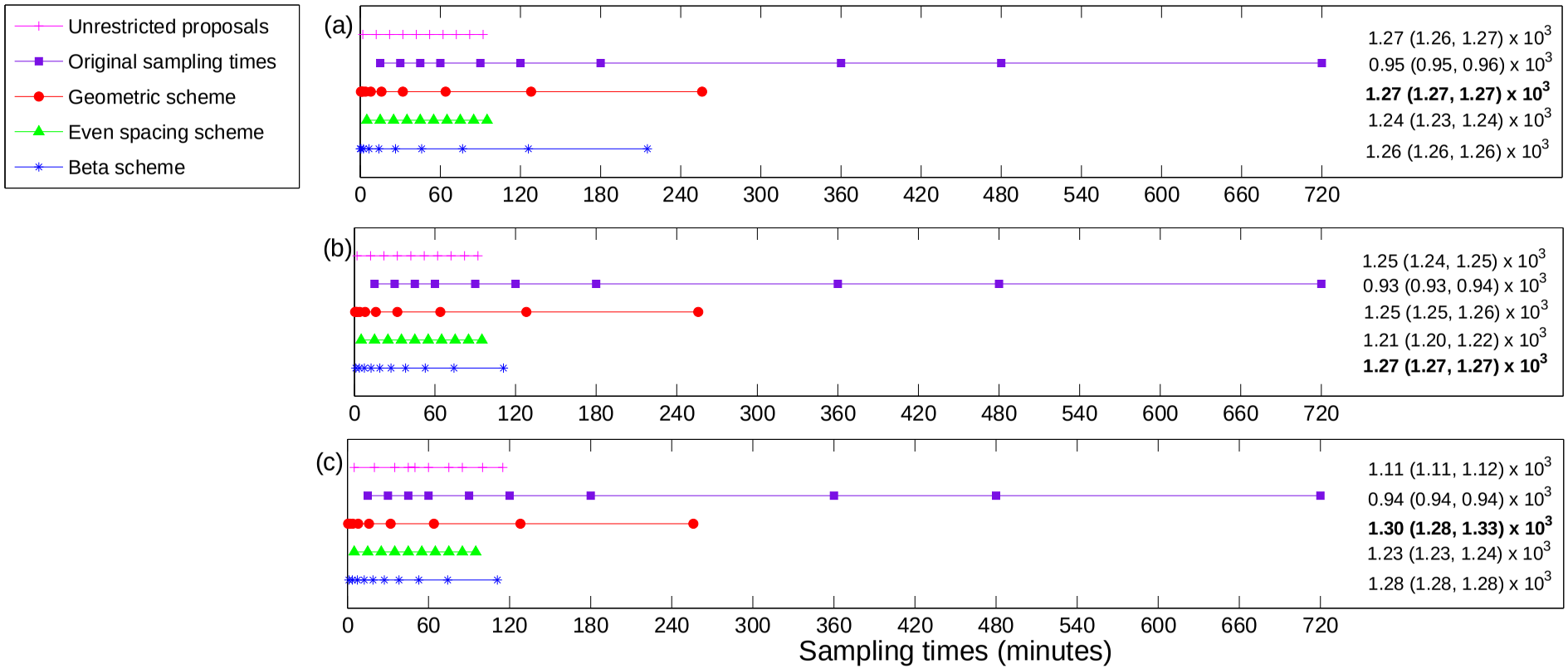


Figure 5.5: The utility is the posterior precision of log AUC: PK sampling times generated under the three lower dimensional parameterisation schemes, as well as the replicate designs and original designs, found using (a) importance sampling, (b) Laplace approximations, and (c) combination of Laplace approximations and importance sampling to estimate the posterior precision of the log AUC utility function.

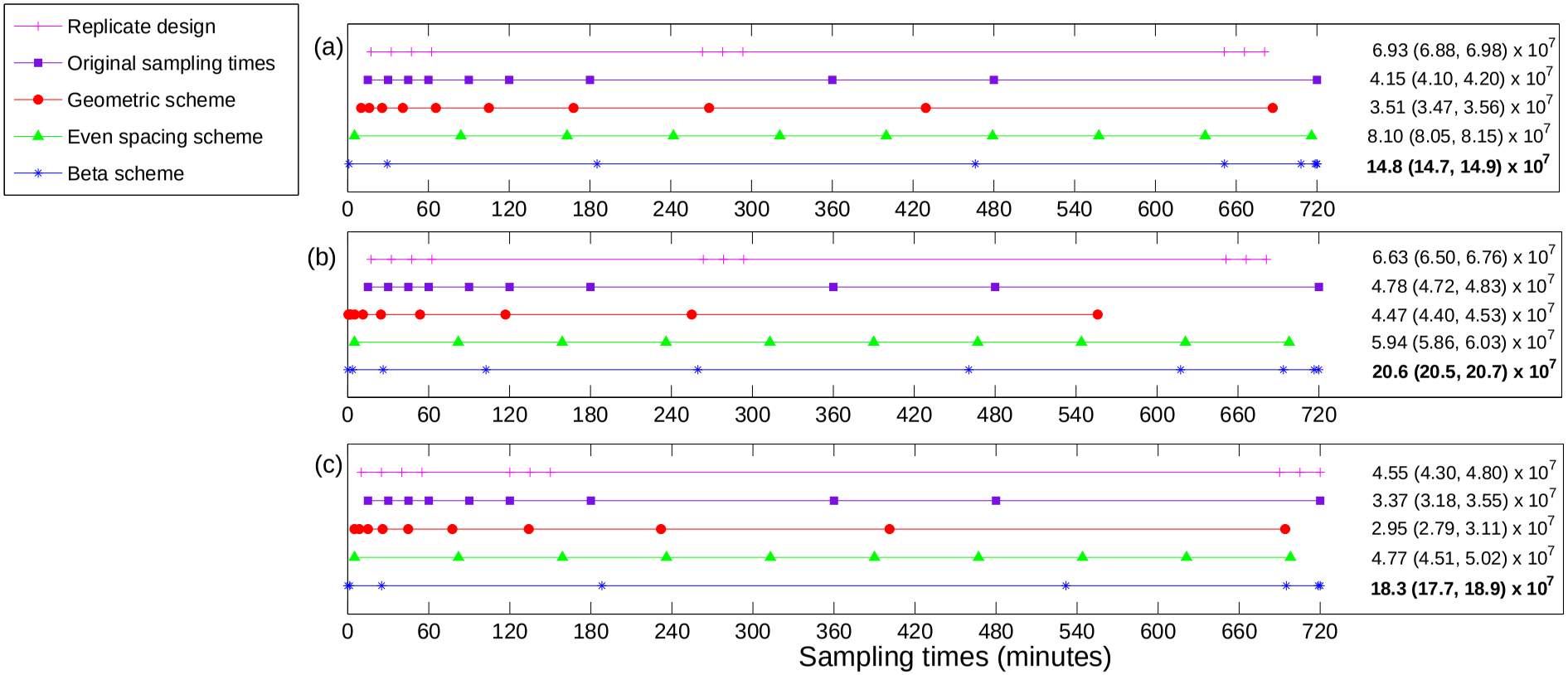


Figure 5.6: The utility is the inverse of the determinant of the posterior variance-covariance matrix: PK sampling times generated under the three lower dimensional parameterisation schemes, as well as the replicate designs and original designs, found using (a) importance sampling, (b) Laplace approximations, and (c) combination of Laplace approximations and importance sampling to estimate the inverse of the determinant of the posterior variance-covariance matrix utility function.

Method for estimating utility function	Utility function	Run time (minutes)[a]
Importance sampling	U_1	59
	U_2	362
	U_3	44
	U_4	85
Laplace approximation	U_1	2
	U_2	5
	U_3	2
	U_4	1
Combined approach	U_1	10
	U_2	91
	U_3	7
	U_4	23

Table 5.1: Run times for the different methods for calculating the utility functions for when the geometric scheme is used to generate the designs. Here U_1 = posterior precision of C_{max}; U_2 = Posterior precision of the difference in C_{max} for ECMO vs non-ECMO; U_3 = posterior precision of log AUC; and U_4 = 1/det(posterior var-cov). [a]These run times are based on MCMC runs of 10 000 iterations.

In general, the estimates of the utility functions were quite similar across the three methods for calculating the utilities (see Figures 5.3 - 5.6). The run times for the different utilities varied significantly across the three methods in most cases, with the Laplace approximation being the fastest, followed by the combined approach and then importance sampling (from the prior). In particular, the run times for the posterior precision of the absolute difference in peak concentration estimates for sheep on ECMO vs sheep not on ECMO utility ($U_2(\mathbf{d}, \mathbf{y})$) were much slower for importance sampling compared to the other methods. This is due to the fact that importance sampling required a larger number of particles to provide a stable estimate of the utility. The importance sampling run time for the Bayesian D-posterior precision utility could have been just as slow (if not slower) if no upperbound on M_p was present.

For the posterior precision of C_{max} and the posterior precision of the difference in C_{max} for ECMO vs non-ECMO utility functions, the designs were quite similar across the three methods for estimating the utility functions. Sampling was mostly focused around the peak concentration (the beta scheme and occasionally the replicate designs also sampled beyond these times). For the posterior precision of the log AUC utility, the designs mostly focused on sampling in areas where the concentration was non-zero, and were quite similar across the three methods for estimating the utilities. However, the geometric scheme covered a greater area of the concentration-time curve and also sampled where the concentrations were approximately zero. The designs which were generated for the inverse of the determinant of the posterior variance-covariance matrix utility function were similar across the three methods and covered the majority of the sampling space. The beta scheme produced the optimal designs for this utility, and the beta distribution was in the shape of a 'bath tub' curve over the design space. The original sampling times which were used by Shekar et al. (2013) were always outperformed by the optimal designs that were found in this paper, for our utility functions of interest, highlighting the importance of these methods for determining an optimal experimental design.

Method	U_{Orig}	(a, b)	R_{U_1}	R_{U_2}	R_{U_3}	R_{U_4}
IS	U_1	$(0.21, 6.55)$	-	0.91 (0.91, 0.92)	0.72 (0.71, 0.72)	0.09 (0.09, 0.10)
	U_2	$(0.3, 6.2)$	0.95 (0.94, 0.95)	-	0.93 (0.93, 0.94)	0.03 (0.03, 0.04)
	U_3	$(0.3, 3)$	0.83 (0.83, 0.84)	0.98 (0.98, 0.99)	-	0.04 (0.04, 0.05)
	U_4	$(0.2, 0.1)$	0.43 (0.43, 0.43)	0.82 (0.82, 0.82)	0.46 (0.45, 0.47)	-
LA	U_1	$(0.2, 3)$	-	0.93 (0.93, 0.94)	1.02 (1.01, 1.02)	0.02 (0.02, 0.02)
	U_2	$(0.21, 6.55)$	1.08 (1.07, 1.08)	-	0.64 (0.63, 0.65)	0.01 (0.01, 0.01)
	U_3	$(0.6, 10)$	0.92 (0.91, 0.93)	0.84 (0.84, 0.85)	-	0.01 (0.01, 0.01)
	U_4	$(0.2, 0.2)$	0.60 (0.60, 0.61)	0.57 (0.56, 0.58)	0.59 (0.59, 0.60)	-
C	U_1	$(0.2, 1.7)$	-	0.92 (0.92, 0.92)	1.03 (1.02, 1.03)	0.31 (0.31, 0.32)
	U_2	$(0.3, 10)$	1.07 (1.06, 1.07)	-	0.90 (0.90, 0.91)	0.29 (0.29, 0.30)
	U_3	$(0.6, 10)$	1.02 (1.02,1.03)	0.91 (0.90, 0.91)	-	0.29 (0.28, 0.30)
	U_4	$(0.1, 0.1)$	0.80 (0.80,0.81)	0.72 (0.72, 0.73)	0.99 (0.98, 0.99)	-

Table 5.2: Comparison of the optimal beta proposal scheme designs for each utility function evaluated at the other utility functions, across the different methods for calculating the utilities. U_{Orig} is the utility function that the beta scheme design originally came from; (a, b) are the values of the shape parameters for the beta scheme; R_{U_1}, R_{U_2}, R_{U_3}, and R_{U_4} are the values of the utilities (with 95% CI in brackets) evaluated at the beta scheme design which came from the utility U_{Orig} divided by the utility evaluated at its own beta scheme design (ratios). Here U_1 = posterior precision of C_{max}; U_2 = Posterior precision of the difference in C_{max} for ECMO vs non-ECMO; U_3 = posterior precision of log AUC; and $U_4 = 1/\det$(posterior var-cov); IS = importance sampling; LA = Laplace approximation; and C = combined approach.

Contour plots of the posterior samples of the beta scheme design variables (a, b) for each of the utility functions and methods of calculating the utilities are displayed in Figure 5.7 in Appendix B.

For each of the utility functions, the beta scheme was often found to perform quite well, across the three methods for calculating the utilities. Therefore, it was decided to perform a comparison of the (optimal) designs generated by the beta schemes across the different utility functions and methods for calculating the utilities, to see how these designs differed. A graphical comparison of the beta scheme designs is displayed in Figure 5.8 in Appendix C.

A quantitative comparison was performed, in which the beta scheme design from one utility function was input into the other three utility functions and their values were calculated. A ratio was then calculated in which these utility function values were compared to the values of those utility functions evaluated at their own beta scheme designs (columns 4-7 in Table 5.2).

The majority of the utility functions performed quite well (as indicated by high ratios) when designs from other utility functions were input. The exception to this was that when the designs from the other three utilities were input into the inverse of the determinant of the posterior variance-covariance matrix utility, low ratio values were obtained for all three methods of estimating the utility. This is not surprising given how different the designs obtained for the inverse of the determinant of the posterior variance-covariance matrix utility were compared to the designs obtained for other utility functions (see Figures 5.3 - 5.6). These results suggest that this utility function is not robust to design objective uncertainty, and if one is interested in precisely estimating all PK model parameters, then one should specifically design the experiment to do so and not rely on other designs.

5.6 Discussion

In this paper we have compared and contrasted three methods for calculating Bayesian utility functions: importance sampling using the prior as the importance distribution; Laplace approximations; and importance sampling using the Laplace approximation to the posterior as the importance distribution. These approaches to calculating the utility functions were incorporated into an MCMC algorithm which searched for the (near) optimal design for a PK study, which required 10 plasma sampling times to be found. Four Bayesian utility functions were used which focused on precisely estimating various PK measures of interest and were functions of the PK parameters.

The optimal designs that were found differed substantially between the utility functions, but were fairly similar between the different methods for calculating the utility functions (for a given utility function). The posterior precision of C_{max}, the posterior precision of the difference in C_{max} for ECMO vs non-ECMO, and the posterior precision of the log AUC utility functions were found to be fairly robust to uncertainty in the design objectives. However, the inverse of the determinant of the posterior variance-covariance matrix utility was not found to be robust to design objective uncertainty. This means that designs that are generated by other utility function should not be used if one is interested in precisely estimating all PK parameters.

The Laplace approximation method was generally found to be the fastest of the three methods. The combined approach was computationally faster than the importance sampling (from the prior) approach since fewer samples of $(\boldsymbol{\theta}, \mathbf{y})$ were required to obtain stable and precise estimates of the utilities. When importance sampling was used to estimate the utility functions, many importance samples were required to obtain reasonable ESS values. Both the Laplace approximations and the 'combined approach' were able to produce similar results to brute force importance sampling, but in a more timely manner. This is of high importance when one is interested in designing experiments which involve large amounts of data to be collected.

As an alternative to importance sampling, which can break down for large amounts of data, we used Laplace approximations to calculate the Bayesian utility functions. The use of each of these methods for approximating the posterior distribution is problem-dependent. Previously, importance sampling from the prior has been used as a gold standard (e.g., Cook et al. (2008); Ryan et al. (2014c)). However, if large amounts of data are involved, then we do not recommend the use of importance sampling from the prior distribution as this method was found to be computationally intensive, due to the large number of particles required to obtain a reasonable ESS. The Laplace approximation approach is useful when large amounts of data are involved, but its suitability depends on whether it is reasonable to assume that the posterior distribution follows a multivariate normal distribution. This could be a reasonable assumption in many design applications where large amounts of data are involved and/or if the priors are reasonably informative. If the Gaussian assumption is not reasonable, then we do not recommend the use of the Laplace approximation for estimating the posterior distribution. The "combined

approach" could be used for a wider variety of design problems, as it corrects for some non-normality in the Laplace approximation, and can be used for large amounts of data since fewer particles are required in the importance sampling to obtain a reasonable ESS (hence reducing the computational burden), due to the fact that the importance distribution in the combined approach is guided by an approximation to the target (posterior distribution). However, for a high degree of non-normality, the combined approach may not be useful. Alternative methods for 'fast' posterior approximation should be investigated, such as sequential Monte Carlo with a Liu West filter (Liu and West (2001)), or adaptive importance sampling (e.g., Kinas (1996); Pennanen and Koivu (2006)).

To ease the computational burden of searching over a large number of design points, we used lower dimensional parameterisations, which reduced the number of design variables to search over from ten to two. We investigated three different schemes that would generate the 10 design points after values of the two design variables had been chosen. For the most part, the designs generated by the lower dimensional schemes were quite different. The schemes were chosen with our PK application in mind, but other functions or transformations may be more suitable for different design problems. The beta proposal scheme was found to be quite flexible in generating the designs, in that a wide variety of designs could be generated from this scheme depending on the values of the shape parameters used, and so it may generally be a good lower dimensional scheme to use for a wide variety of design problems. Additional flexibility could be obtained by including another design variable in the parameterisation of the beta proposal scheme that determines the optimal percentiles of the beta distribution to use. If one is unsure what lower dimensional scheme may be most appropriate for their design problem, we recommend running several different parameterisations in parallel on different CPUs (as we have done) and choosing the scheme that generates the design with the highest utility value.

For all of the utility functions, we were able to find an alternative design that produced higher utility function values than the design that was used by Shekhar et al. (2013). This suggests that for the next sheep in the experiment the design could be adjusted, as per the results in this paper, depending on the design objective. This also highlights that substantial gains can be achieved if one has the flexibility of being able to adapt the design for each new subject in light of the information obtained from previous subjects. Also, the majority of the utility functions preferred early sampling times, which is practically useful for our motivating study as it would reduce the duration of the study and hopefully study costs.

The utility functions that we used in this study focused on precisely estimating one (or two) PK measure(s) of interest. Future studies may wish to investigate the use of compound design criteria which could focus on designing for the precise estimation of several PK measures of interest.

A fixed, and somewhat large number of sampling times were employed for the examples used in this work, so that the performance of importance sampling and Laplace approximations could be compared for approximating the posterior distribution when large

amounts of data are involved. Investigation of the optimal number of sampling times was outside the scope of this work. The number of sampling times used in this study may not be optimal, particularly if there are cost constraints involved. It is likely that the increase in the expected utility value would plateau after a certain number of observations. We are currently conducting studies which investigate the optimal number of subjects and samples per subjects in population PK studies, where there are cost constraints involved.

Acknowledgments

E.G. Ryan was supported by an APA(I) Scholarship which came from an ARC Linkage Grant with Roche Palo Alto (LP0991602). The work of A.N. Pettitt was also supported by an ARC Discovery Project (DP110100159).

The authors would like to thank Drs Helen Thompson and James McGree from Queensland University of Technology for their helpful comments throughout the process of writing this paper.

Supplementary Materials - Appendices

Appendix A - MCMC algorithm for Bayesian optimal design (Section 5.4)

Algorithm 5.1: MCMC algorithm for Bayesian optimal design

1 Initialise - set an initial design $\mathbf{d}^{(1)}$, simulate $(\boldsymbol{\theta}_j, \mathbf{y}_j)$ from $p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{d}^{(1)}) = p(\boldsymbol{\theta})p(\mathbf{y}|\mathbf{d}^{(1)}, \boldsymbol{\theta})$ for $j = 1, ..., J$

2 Compute $U^{(1)} = \prod_{j=1}^{J} U(\mathbf{d}^{(1)}, \boldsymbol{\theta}_j, \mathbf{y}_j)$

3 **for** *i = 1 to iters* **do**

4 Generate a candidate design $\tilde{\mathbf{d}} \sim q(\cdot|\mathbf{d}^i)$, propose $(\tilde{\boldsymbol{\theta}}_j, \tilde{\mathbf{y}}_j) \sim p(\boldsymbol{\theta}, \mathbf{y}|\tilde{\mathbf{d}}) = p(\boldsymbol{\theta})p(\mathbf{y}|\tilde{\mathbf{d}}, \boldsymbol{\theta})$ for $j = 1, ..., J$

5 If $\tilde{\mathbf{d}}$ is not within the design space then reject the proposal and go to line 9

6 Compute $\tilde{U} = \prod_{j=1}^{J} U(\tilde{\mathbf{d}}, \tilde{\boldsymbol{\theta}}_j, \tilde{\mathbf{y}}_j)$

7 Calculate the MH acceptance probability, $a = \min(1, A)$ where

$$A = \frac{\tilde{U} \times q(\mathbf{d}^{(i)}|\tilde{\mathbf{d}})}{U^{(i)} \times q(\tilde{\mathbf{d}}|\mathbf{d}^{(i)})}$$

Here $U^{(i)}$ and $\mathbf{d}^{(i)}$ are the current utility and design point values, respectively, and $\tilde{U}$ and $\tilde{\mathbf{d}}$ are the proposed utility and design point values, respectively.

8 Set $(\mathbf{d}^{(i+1)}, U^{(i+1)}) = (\tilde{\mathbf{d}}, \tilde{U})$ with probability a, and

9 $(\mathbf{d}^{(i+1)}, U^{(i+1)}) = (\mathbf{d}^{(i)}, U^{(i)})$ with probability $1 - a$.

Appendix B - Expected utility surfaces for the various utility functions used in the ECMO PK example (Section 5.5)

To determine the optimal designs, we searched for the bivariate mode of the multivariate normal kernel smoothing density estimates of the design variables (see Cook et al. (2008); Drovandi and Pettitt (2013)). The contour plots (Figure 5.7) were used to examine the bivariate modes and to determine whether the MCMC algorithms had converged.

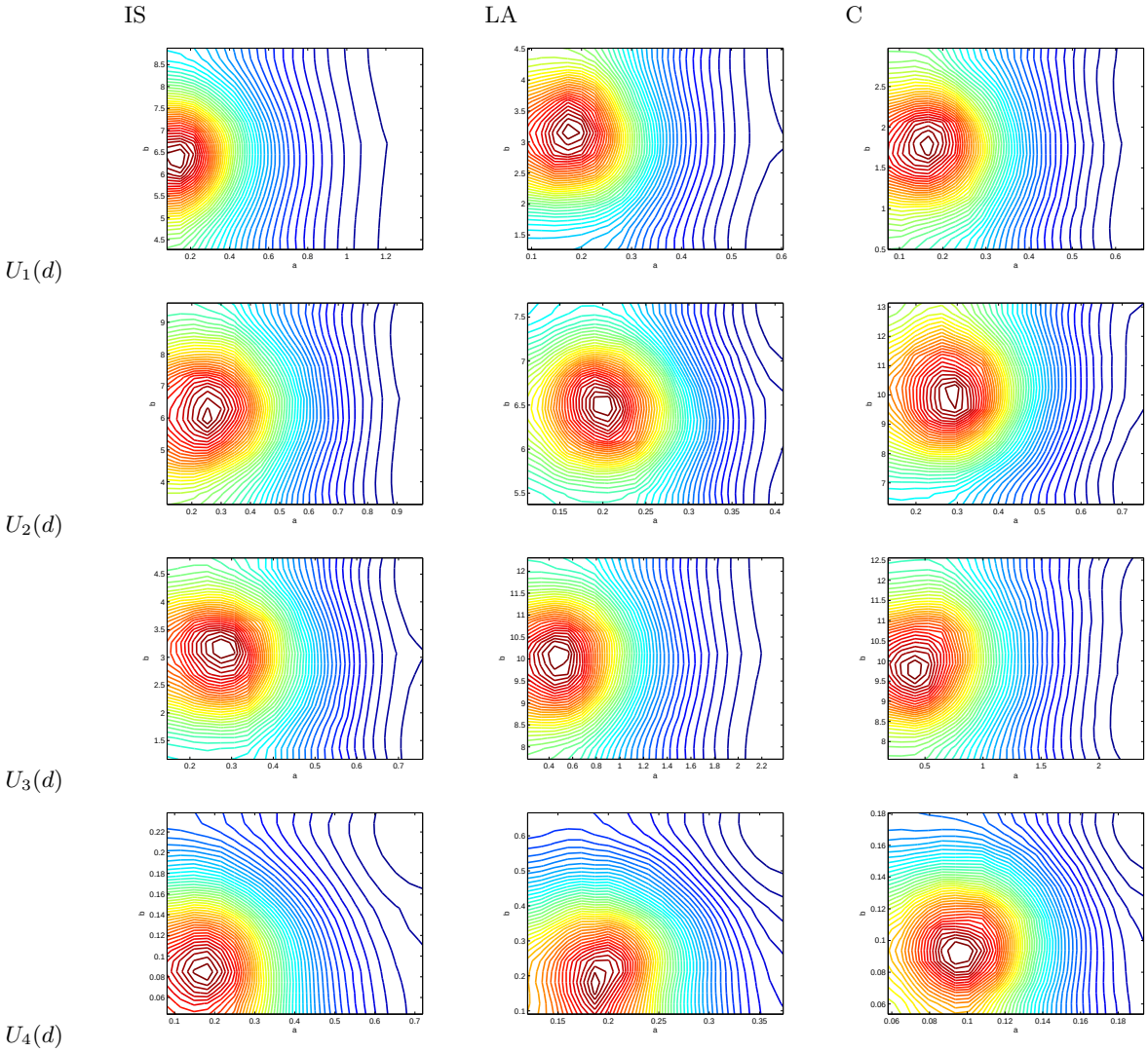


Figure 5.7: Contour plots of the expected utility surface for the beta scheme, for the different utility functions (rows) and three methods for calculating the utility functions (columns). Here U_1 = posterior precision of C_{max}; U_2 = Posterior precision of the difference in C_{max} for ECMO vs non-ECMO; U_3 = posterior precision of log AUC; and $U_4 = 1/\det$(posterior var-cov); IS = importance sampling; LA = Laplace approximation; and C = combined approach.

Appendix C - Comparison of designs from the beta proposal scheme for the ECMO PK example (Section 5.5)

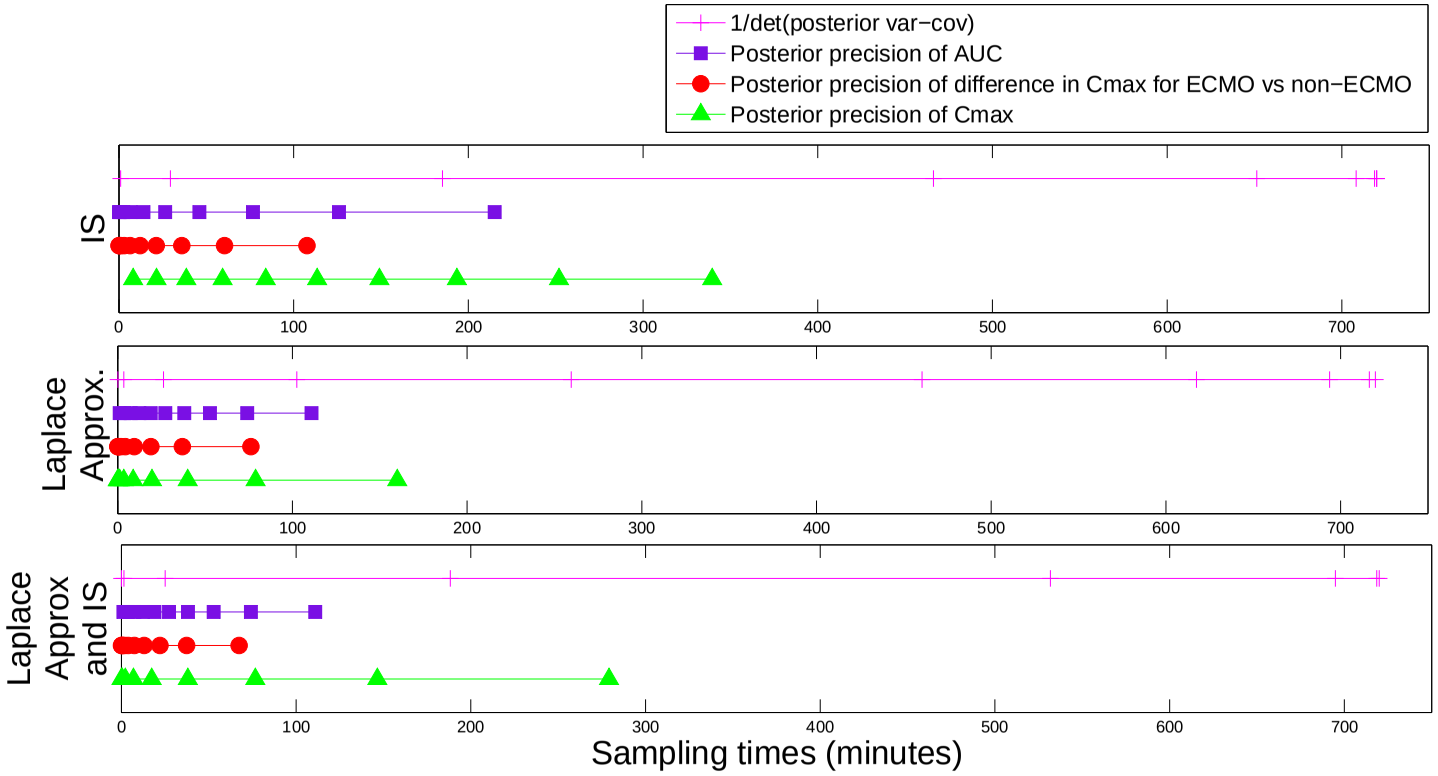


Figure 5.8: Comparison of the designs which arose from the beta proposal scheme from the various utility functions across the three different methods for calculating the utilities.

From Figure 5.8, it can be seen that the designs that arose from the beta schemes varied widely across the utility functions, and were somewhat different across the methods for calculating the utility functions.

Chapter 6

Summary and Discussion of Future Directions of Research

In this chapter we summarise the results and novel contributions of the thesis, and suggest areas for future research.

6.1 Summary

The aim of this thesis is to extend upon and advance methods for Bayesian experimental design, particularly for complicated static design problems. The specific aims of this thesis, as described in Chapter 1, are:

(a) To extend existing MCMC algorithms to solve static design problems which require a large number of design points to be found through the development of proposal distributions for the design points that rely on lower dimensional parameterisations of the design points;
(b) To enable fully Bayesian static designs to be found for the collection of data that is modelled by NLMEMs;
(c) To find efficient methods for estimating the posterior distribution for use in Bayesian utility function calculation;
(d) To develop novel design criteria, based on the posterior distribution of some parameter of interest, which may be used to assist in the selection of optimal sampling times;
(e) To apply the above-mentioned methods to case study data, to improve upon the design of existing pharmacokinetic studies, demonstrate the effectiveness of the methods, and evaluate their practical benefits for the pharmaceutical and drug testing industries.

In Chapter 3 we proposed methods which may be used to solve design problems in which one is interested in finding a large number of (near) optimal design points (for a small number of different design variables). The approach involved the use of lower dimensional parameterisations that consisted of a few design variables which generated a large number of design points. Using this approach, one simply has to search over a few design variables (instead of searching over a large number of design variables), which provides substantial computational savings and a more accurate determination of the mode of the utility surface. This approach was incorporated into existing MCMC algorithms (e.g., Müller (1999)) which had previously been used to find optimal Bayesian designs for low

dimensional design problems (up to four design variables). Our methodology was demonstrated on a number of different applications, including the selection of sampling times for pharmacokinetic and heat transfer studies, and involved nonlinear models.

In Chapter 4 we extended the work of Müller (1999) and Stroud et al. (2001) to find fully Bayesian static designs for NLMEMs. This involved the use of simulation-based optimal design methods to search over both continuous and discrete design spaces for a number of different design variables so that optimal population designs for NLMEMs could be found. Although Bayesian inference has commonly been performed on NLMEMs, there has been a lack of research into performing Bayesian optimal design for NLMEMs that require searches to be performed over several design variables. Previous approaches have also found optimal Bayesian designs for NLMEMs by searching over a finite set of designs, and to our knowledge, no other studies have searched over a continuous design space to find optimal Bayesian static designs for NLMEMs. Our methods were used to design for a horse population pharmacokinetic study, in which the design problem was to determine the optimal number of subjects and samples per subject, as well as the (near) optimal urine sampling times, so that the population pharmacokinetic parameters could be precisely estimated, subject to cost constraints.

Utility functions in Bayesian experimental design are based on the posterior distribution. When the posterior is found by simulation, it must be sampled from for each future data set drawn from the prior predictive distribution, and so many thousands of posterior distributions are often required. In Chapter 5 we compared and contrasted the use of importance sampling and Laplace approximations to rapidly approximate the posterior distribution for use in Bayesian utility function calculation. These methodologies were used to solve design problems which involved a somewhat large amount of data. Importance sampling from the prior distribution tends to break down when there is a reasonable number of experimental observations, and so Laplace approximations were used to overcome this. We also considered using the Laplace approximation to form the importance distribution to obtain a more efficient importance distribution than the prior, and to our knowledge, this has not been previously implemented. The methodology was motivated by a pharmacokinetic study which investigated the effect of extracorporeal membrane oxygenation on the pharmacokinetics of antibiotics in sheep. The design problem was to find 10 near optimal plasma sampling times which produced precise estimates of pharmacokinetic model parameters/measures of interest. We considered several different utility functions, some of which were novel, which involved the posterior distribution of parameter functions.

6.2 Future Research

We believe the future of Bayesian experimental design lies in: (1) developing and implementing fast methods for approximating the posterior distribution for use in Bayesian utility functions, and fast computation of the Bayesian utility functions, as these are the most computationally intensive components of Bayesian experimental design; and (2) finding solutions to complex Bayesian experimental design problems, such as problems in

which the likelihood is intractable or computationally prohibitive to calculate, or problems with a large number of design points and design dimensions.

6.2.1 Fast Algorithms for Bayesian Experimental Design

MCMC and importance sampling have been found to be computationally intensive to perform at each iteration of the optimisation algorithm that searches over the space $(\mathbf{d}, \boldsymbol{\theta}, \mathbf{y})$, due to the large number of samples that are required to ensure that the Bayesian utility is well estimated. In particular, importance sampling from the prior performs poorly when large amounts of data are involved due to a low ESS (Ryan et al. (2014a)). Adaptive importance sampling (e.g., Kinas (1996); Pennanen and Koivu (2006)) may provide a faster method for approximating the posterior distributions, but is yet to be explored for Bayesian experimental design.

Laplace approximations and numerical quadrature have been found to be fast alternatives for approximating the posterior distribution in Bayesian design, and can be used when large amounts of data are involved, but rely on the assumption that the posterior distribution follows a multivariate normal distribution and also suffer from the "curse of dimensionality". INLA can also provide a fast method for approximating the posterior distribution, but has not been used for Bayesian experimental design. VB methods are a fast method for facilitating approximate inference for intractable posterior distributions, but are yet to be used in a Bayesian experimental design context.

Drovandi and Pettitt (2013) and Hainy et al. (2013) have explored the use of ABC rejection (see Beaumont et al. (2002)) within an MCMC framework to approximate the posterior distributions for Bayesian utility functions for design problems in which the likelihood function is intractable. Further use of ABC methods for posterior distribution approximation should be explored in the experimental design context.

A few studies have investigated the use of SMC for approximating the necessary quantities for Bayesian utility functions (e.g., Drovandi et al. (2013)), but its use has been limited. Future studies should focus on extending previous approaches to allow for more complicated design problems. SMC with a Liu West filter (Liu and West (2001)) could offer a fast method for posterior approximation for Bayesian design problems.

Computational burden is a major obstacle in all Bayesian design problems for complex models and must be overcome so that designs can be obtained efficiently and in real time, and to broaden the applicability of Bayesian design methodology by making it more accessible to practitioners, scientists and industry. This may be achieved through algorithmic developments and the exploitation of current parallel computing technology (such as graphics processing units or GPUs). Indeed, new parallel architectures are becoming increasingly available to individual researchers, and will have a significant impact on Bayesian experimental design. In order to take advantage of this increased power, computational problems and approaches should be adapted from the current serial processing paradigm to one that optimises algorithms for parallel processing. To our knowledge,

there is no published, peer reviewed research on the use of GPUs in the derivation of a Bayesian experimental design.

6.2.2 Finding Optimal Designs for Complex Models

The future of Bayesian experimental design also lies in solving complex or nonstandard problems, such as problems in which the likelihood is intractable or computationally prohibitive to evaluate, problems where the observed data likelihood cannot be evaluated analytically, or problems with a large number of design points. Whilst sophisticated inference techniques are available for Bayesian data analysis for complex data models, corresponding methodology for deriving Bayesian experimental designs is severely lacking, and it is important that the methods for inference are complemented with appropriate experimental design methodologies that enable more informative data to be collected in a more timely manner. Use of parallel computing technology may be required to ease the computational burden of finding optimal Bayesian experimental designs for complex models (such as mixed effects models).

Fully Bayesian experimental designs for NLMEMs are largely unexplored. Most of the current work has focused on evaluating Bayesian utility functions for a fixed set of discrete designs (e.g., Han and Chaloner (2004); Palmer and Müller (1998)) and selecting the design that produces the highest utility value (i.e., no search over a continuous design space is performed). In Chapter 4, we extended this by searching over a continuous design space to determine (near) optimal sampling times for a horse population pharmacokinetic study. Kim et al. (2013) find optimal sequential designs for population studies. Further work on using SMC algorithms (Chopin (2002)) to search for optimal designs for mixed effects models in the presence of model uncertainty is currently being conducted, so that solutions to real-world design problems can be found. The main difficulty in finding solutions to experimental design problems in which the data is modelled by mixed effects models is that the observed data likelihood is unavailable in closed form for all but the simplest examples.

6.2.3 Finding Optimal Designs for a Large Number of Design Variables

Better search algorithms are also required to find static designs. Many of the search algorithms for obtaining optimal designs (e.g., Müller (1999); Amzal et al. (2006)) are restricted to a small number of design variables (≤ 4), as these algorithms are computationally prohibitive for a large number of design variables (e.g., Bielza et al. (1999); Müller (1999); Stroud et al. (2001); Cook et al. (2008)). MCMC is good at estimating marginal distributions derived from (high dimensional) joint distributions but not good at estimating joint distributions; a case of suffering from the "curse of dimensionality". Therefore estimating a joint multivariable mode is hard. Finding a joint mode analytically is easy. It is when the joint density is approximated by Monte Carlo samples it becomes a hard problem.

In Chapter 3 we proposed the use of lower dimensional parameterisations to enable near optimal designs to be found for problems that require a large number of design points.

The lower dimensional parameterisations consisted of a few design variables, which were optimised, and were then input into various functions to generate multiple design points. This was found to have substantial computational savings, and it was much easier to obtain the multivariate mode for a few design variables than for a large number of design variables. We found that the beta proposal scheme, where the designs come from the (evenly-spaced) percentiles of a beta distribution, gave quite flexible designs, and so this function may be appropriate in many situations to generate a large number of design points. One could also extend the beta proposal scheme to propose from a generalised beta distribution (e.g., Sepanski and Kong (2007)), which may offer further flexibility in constructing the designs. One could also include another design variable in the parameterisation of the beta proposal scheme that determines the optimal percentiles of the beta distribution to use, e.g., percentile $= 100((\frac{i}{n})^{\alpha})$ where α is an additional design variable to search over.

However, designs found using the methods proposed in Chapter 3 are not optimal but *near* optimal, which is a compromise of the computational savings achieved. The approach is only useful for design variables (e.g., sampling times/locations) that require multiple measures to be taken at specific points that are separated from one another in the design space. This approach does not overcome the problem of having a large number of different types of design variables (e.g., temperatures, pressures), and further research needs to be conducted for solving this design problem.

References

Alderman, J., Wolkow, R., Chung, M., and Johnston, H. (1998). Sertraline treatment of children and adolescents with obsessive-compulsive disorder or depression: Pharmacokinetics, tolerability, and efficacy. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37:386–394.

Amzal, B., Bois, F., Parent, E., and Robert, C. P. (2006). Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association*, 101(474):773–785.

Anscombe, F. (1968). Sequential medical trials. *Journal of the American Statistical Association*, 58:365–383.

Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, 58:59–76.

Atkinson, A. C. (2008). Dt-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, 138:56–64.

Atkinson, A. C., Chaloner, K., Herzberg, A., and Juritz, J. (1993). Optimum experimental designs for properties of a compartmenal model. *Biometrics*, 49(2):325–337.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford University Press, New York.

Beal, S. and Sheiner, L. (2002). NONMEM user's guide. Technical report, University of California, San Francisco, NONMEN Project group.

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035.

Bengtsson, T., Bickel, P., and Li, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability and Statistics: Essays in Honor of David A. Freedman*. Institute of Mathematical Statistics.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 2nd edition.

Bernardo, J. M. (1979). Expected information as expected utility. *Annals of Statistics*, 7(3):686–690.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Wiley and Sons, New York.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. John Wiley & Sons, 2nd edition.

Berry, D. (1993). A case for Bayesianism in clinical trials (with discussion). *Statistics in Medicine*, 12:1377–1404.

Berry, D. (2006). Bayesian clinical trials. *Nature Reviews Drug Discovery*, 5:27–36.

Berry, D., Müller, P., Grieve, A., Smith, M., Parke, T., Blazek, R., Mitchard, N., and Krams, M. (2000). *Case Studies in Bayesian Statistics*, chapter Adaptive bayesian designs for dose-ranging drug trials. Springer.

Berry, D. A. and Ho, C.-H. (1988). One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics*, 44:219–227.

Bielza, C., Müller, P., and Insua, D. R. (1999). Decision analysis by augmented probability simulation. *Management Science*, 45(7):995–1007.

Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011). Response-adaptive dose-finding under model uncertainty. *Annals of Applied Statistics*, 5(2B):1611–1631.

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C., König, F., Krams, M., Liu, Q., Neuenschwander, B., Parke, T., and Pinheiro, J. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials (with discussion). *Journal of Biopharmaceutical Statistics*, 17:965–995.

Borth, D. M. (1975). A total entropy criterion for the dual problem of model discrimination and parameter estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37:77–87.

Box, G. E. P. and Hill, W. J. (1967). Discrimination among mechanistic models. *Technometrics*, 9:57–71.

Brockwell, A. E. and Kadane, J. B. (2003). A gridding method for Bayesian sequential decision problems. *Journal of Computational and Graphical Statistics*, 12(3):566–584.

Brogan, T., Thiagarajan, R., Rycus, P., Bartlett, R., and Bratton, S. (2009). Extracorporeal membrane oxygenation in adults with severe respiratory failure: a multi-center database. *Intensive Care Medicine*, 35:2105–2114.

Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. (2004). Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13:907–929.

Carlin, B., Kadane, J., and Gelfand, A. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics*, 54(3):964–975.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., and Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22(4):887–905.

Chaloner, K. (1984). Optimal bayesian experimental designs for linear models. *Annals of Statistics*, 12:283–300.

Chaloner, K. (1987). An approach to design for generalised linear models. In *Proceedings of the Workshop on Model-oriented data analysis, Wartburg. Lecture Notes in Economics and Mathematical Systems.*, Berlin. Springer.

Chaloner, K. and Larntz, K. (1989). Optimal bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10:273–304.

Chang, Y., van Hal, S., Spencer, P., Gosbell, I., and Collett, P. (2010). Comparison of adult patients hospitalised with pandemic (H1N1) 2009 influenza and seasonal influenza during the "PROTECT" phase of the pandemic response. *Medical Journal of Australia*, 192:90–93.

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–552.

Christen, J., Müller, P., Wathen, K., and Wolf, J. (2004). Bayesian randomized clinical trials: A decision-theoretic sequential design. *Canadian Journal of Statistics*, 32(4):387–402.

Clyde, M. (2001). Experimental design: A Bayesian perspective. *International Encyclopedia of the Social and Behavioural Sciences*, 8:5075–5081.

Clyde, M. and Chaloner, K. (1996). The equivalence of constrained and weighted designs in multiple objective design problems. *Journal of the American Statistical Association*, 91:1236–1244.

Clyde, M. A., Müller, P., and Parmigiani, G. (1996). Exploring expected utility surfaces by Markov Chains. Technical report, Duke University.

Cook, A., Gibson, G., and Gilligan, C. (2008). Optimal observation times in experimental epidemic processes. *Biometrics*, 64(3):860–868.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. (1987). Minimizing multimodal functions of continuous variables with the simulated annealing algorithm. *ACM Transactions on Mathematical Software*, 13:262–280.

D'Argenio, D. (1990). Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Mathematical Biosciences*, 99(1):105–118.

D'Argenio, D. Z. (1981). Optimal sampling times for pharmacokinetic experiments. *Journal of Pharmacokinetics and Pharmacodynamics*, 9:739–756.

DasGupta, A. (1995). Review of optimal Bayes designs. Technical report, Purdue University.

DasGupta, A., Mukhopadhyay, S., and Studden, W. (1992). Compromise designs in heteroscedastic linear models. *Journal of Statistical Planning and Inference*, 32:363–384.

DasGupta, A. and Studden, W. (1991). Robust bayes designs in normal linear models. *Annals of Statistics*, 19:1244–1256.

Davidian, M. and Gallant, R. (1993). The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, 80:475–488.

Davidian, M. and Giltinan, D. M. (2003). Nonlinear models for repeated measures data. In Duan, S. and Reise, N., editors, *Methodological advances issues and applications*, pages 1–24. Chapman & Hall, London.

Davies, A., Jones, D., Bailey, M., Beca, J., Bellomo, R., Blackwell, N., Forrest, P., Gattas, D., Granger, E., Herkes, R., Jackson, A., McGuinness, S., Nair, P., Pellegrino, V., Pettila, V., Plunkett, B., Pye, R., Torzillo, P., Webb, S., Wilson, M., and Ziegenfuss, M. (2009). Extracorporeal membrane oxygenation for 2009 influenza A(H1N1) acute respiratory distress syndrome. *Journal of the American Medical Association*, 302:1888–1895.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGrawHill, New York.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

Dette, H. (1990). A generalization of D-and D1-optimal designs in polynomial regression. *The Annals of Statistics*, 18:1784–1804.

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. (2008). Optimal designs for dose finding studies. *Journal of the American Statistical Association*, 103:1225–1237.

Diggle, P., Liang, K., and Zeger, S. (1994). *Analysis of Longitudinal Data*. Clarendon Press, Oxford.

Diggle, P. and Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33(1):53–64.

Ding, M., Rosner, G. L., and Müller, P. (2008). Bayesian optimal design for phase II screening trials. *Biometrics*, 64:886–894.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Dragalin, V., Bornkamp, B., Bretz, F., Miller, F., Padmanabhan, S., Patel, N., Perevozskaya, I., Pinheiro, J., and Smith, J. (2010). A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research*, 2:487–512.

Dragalin, V., Hsuan, F., and Padmanabhan, S. (2007). Adaptive designs for dose-finding studies based on sigmoid Emax model. *Journal of Biopharmaceutical Statistics*, 17:1051–1070.

Dror, H. and Steinberg, D. (2008). Sequential experimental designs for generalized linear models. *Journal of the American Statistical Association*, 103(481):288–298.

Drovandi, C., McGree, J., and Pettitt, A. (2014). A sequential Monte Carlo algorithm to incorporate model uncertainty in Bayesian sequential design. *Journal of Computational and Graphical Statistics*, 23(1):3–24.

Drovandi, C. C., McGree, J. M., and Pettitt, A. N. (2013). Sequential Monte Carlo for Bayesian sequential design. *Computational Statistics and Data Analysis*, 57(1):320 – 335.

Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.

Drovandi, C. C. and Pettitt, A. N. (2013). Bayesian experimental design for models with intractable likelihoods. *Biometrics*, 69(4):937–948.

Drovandi, C. C., Pettitt, A. N., and Faddy, M. J. (2011). Approximate Bayesian computation using indirect inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(3):503–524.

Duffull, S., Waterhouse, T., and Eccleston, J. (2005). Some considerations on the design of population pharmacokinetic studies. *Journal of Pharmacokinetics and Pharmacodynamics*, 32:441–457.

Duffull, S. B., Graham, G., Mengersen, K., and Eccleston, J. (2012). Evaluation of the preposterior distribution of optimized sampling times for the design of pharmacokinetic studies. *Journal of Biopharmaceutical Statistics*, 22(1):16–29.

El-Krunz, S. and Studden, W. (1991). Bayesian optimal designs for linear regression models. *Annals of Statistics*, 19:2183–2208.

Etzione, R. and Kadane, J. B. (1993). Optimal experimental design for another's analysis. *Journal of the American Statistical Association*, 88:1401–1411.

Federov, V. and Hackl, P. (1997). *Model-oriented Design of Experiments*. Springer-Verlag, Berlin.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Friel, N. and Pettitt, A. N. (2008). Marginal likelihood estimation via power posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):589–607.

Geweke, J. (1989). Bayesian inference in Econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339.

Gibaldi, M. and Perrier, D. (1982). *Pharmacokinetics*. Dekker, New York.

Gotwalt, C., Jones, B., and Steinberg, D. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*, 51:88–95.

Hainy, M., Müller, W., and Wagner, H. (2013). Likelihood-free simulation-based optimal design. Technical report, Johannes Kepler University, Linz.

Han, C. and Chaloner, K. (2004). Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics*, 60:25–33.

Hashimoto, Y. and Sheiner, L. (1991). Designs for population pharmacodynamics: Value of pharmacokinetic data and population analysis. *Pharmacometrics*, 19(3):333–353.

Hiemenz, J., Cagnoni, P., Simpson, D., Devine, S., Chao, N., Keirns, J., Lau, W., Facklam, D., and Buell, D. (2005). Pharmacokinetic and maximum tolerated dose study of micafungin in combination with fluconazole versus fluconazole alone for prophylaxis of fungal infections in adult patients undergoing a bone marrow or peripheral stem cell transplant. *Antimicrobial Agents and Chemotherapy*, 49(4):1331–1336.

Hill, W., Hunter, W., and Wichern, D. (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics*, 10(1):145–160.

Hill, W. J. (1978). A review of experimental design procedures for regression model discrimination. *Technometerics*, 20:15–21.

Ho, Y. C. and Cao, X. (1983). Perturbation analysis and optimization of queueing networks. *Journal of Optimization Theory and Applications*, 40:559–582.

Huan, X. and Marzouk, Y. M. (2012). Gradient-based stochastic optimization methods in bayesian experimental design. Technical report, Massachusetts Institute of Technology, Cambridge.

Huan, X. and Marzouk, Y. M. (2013). Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317.

Hui, D., Lee, N., and Chan, P. (2010). Clinical management of pandemic 2009 influenza A(H1N1) infection. *Chest*, 137(4):916–25.

Johansen, A., Doucet, A., and Davy, M. (2006). Maximum likelihood parameter estimation for maximum likelihood models using sequential Monte Carlo methods. In *ICASSP*.

Jonsson, E., Wade, J., and Karlsson, M. (1996). Comparison of some practical sampling strategies for population pharmacokinetic studies. *Journal of Pharmacokinetics and Biopharmaceutics*, 24(2):245–263.

Kadane, J. B. (1996). *Bayesian methods and ethics in clinical trial design*. John Wiley & Sons.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B*, 21:272–304.

Kiefer, J. (1961). Optimum designs in regression problems. II. *Annals of Mathematical Statistics*, 32(2):298–325.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, 2(5):849–1063.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs on regression problems. *Annals of Mathematical Statistics*, 30:271–94.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 14:363–366.

Kiefer, J. and Wolfowitz, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. *Annals of Mathematical Statistics*, 36(6):1627–1655.

Kim, W., Pitt, M. A., Lu, Z.-L., Steyvers, M., and Myung, J. I. (2013). A hierarchical adaptive approach to optimal experimental design. Technical report, Ohio State University.

Kinas, P. (1996). Bayesian fishery stock assessment and decision making using adaptive importance sampling. *Canadian Journal of Fisheries and Aquatic Sciences*, 53:414–423.

Kück, H., de Freitas, N., and Doucet, A. (2006). Smc samplers for Bayesian optimal nonlinear design. Technical report, University of British Columbia.

Kujala, J. V. and Lukka, T. J. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology*, 50(4):369–389.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Kuo, L., Soyer, R., and Wang, F. (1999). *Bayesian Statistics VI*, chapter Optimal design for quantal bioassay via Monte Carlo methods, pages 795–802. Oxford University Press, New York.

Lesmes, L., Jeon, S.-T., Lu, Z.-L., and Dosher, B. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method. *Vision Research*, 46:3160–3176.

Lewi, J., Butera, R., and Paninski, L. (2009). Sequential optimal design of neurophysiology experiments. *Neural Computation*, 21:619–687.

Lewis, R. and Berry, D. A. (1994). Group sequential clinical trials: A classical evaluation of Bayesian decision-theoretic designs. *Journal of the American Statistical Association*, 89:1528–1534.

Liepe, J., Filippi, S., Komorowski, M., and Stumpf, M. P. H. (2013). Maximising the information content of experiments in systems biology. *PLoS Computational Biology*, 9(1):e1002888. doi:10.1371/journal.pcbi.1002888.

Lindley, D. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27:986–1005.

Lindley, D. (1968). The choice of variables in multiple regression. *Journal of the Royal Statistical Society Series B*, 30:31–53.

Lindley, D. (1972). *Bayesian Statistics - A Review*. SIAM, Philadelphia.

Lindstrom, M. and Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics*, 46(3):673–687.

Liu, G., Rosenberger, W. F., and Haines, L. M. (2009). Sequential designs for ordinal phase I clinical trials. *Biometrical Journal*, 51(2):335–347.

Liu, J. and West, M. (2001). *Sequential Monte Carlo Methods in Practice*, chapter 10: Combined Parameter and State Estimation in Simulation-Based Filtering. Springer Verlag New York.

Loeppky, J., Moore, L., and Williams, B. (2010). Batch sequential designs for computer experiments. *Journal of Statistical Planning and Inference*, 140:1452–1464.

Long, Q., Scavino, M., Tempone, R., and Wang, S. (2013). Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Computer Methods in Applied Mechanics and Engineering*, 259:24–39.

López-Fidalgo, J., Tommani, C., and Trandafir, P. (2007). An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society:Series B (Statistical Methodology)*, 69:231–242.

Loredo, T. (2004). Bayesian adaptive exploration. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 23rd International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, pages 330–346.

Lunn, D. J., Best, N., Thomas, A., Wakefield, J., and Spiegelhalter, D. (2002). Bayesian analysis of population PK/PD models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics*, 29:271–307.

Maryak, J. and Chin, D. (2004). Global random optimization by simultaneous perturbation stochastic approximation. *Johns Hopkins APL Technical Digest*, 25(2):91–100.

Masoumi, S., Duever, T. A., and Reilly, P. M. (2013). Sequential Markov chain Monte Carlo (MCMC) model discrimination. *The Canadian Journal of Chemical Engineering*, 91(5):862–869.

McGree, J., Drovandi, C. C., and Pettitt, A. N. (2012a). A sequential Monte Carlo approach to derive sampling times and windows for population pharmacokinetic studies. *Journal of Pharmacokinetics and Pharmacodynamics*, 39(5):519–526.

McGree, J., Drovandi, C. C., and Pettitt, A. N. (2012b). A sequential Monte Carlo approach to the sequential design for discriminating between rival continuous data models. Technical report, Queensland University of Technology.

McGree, J., Drovandi, C. C., Thompson, H., Eccleston, J., Duffull, S., Mengersen, K., Pettitt, A. N., and Goggin, T. (2012c). Adaptive Bayesian compound designs for dose finding studies. *Journal of Statistical Planning and Inference*, 142(6):1480–1492.

McGree, J., Noble, G., Schneiders, F., Dunstan, A., McKinney, A., Boston, R., and Sillence, M. (2012d). A Bayesian approach for estimating detection times in horses: exploring the pharmacokinetics of a urinary acepromazine metabolite. *Journal of Veterinary Pharmacology and Therapeutics*.

Mentré, F., Mallet, A., and Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika*, 84(2):429–442.

Meyer, R. and Nachtsheim, C. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometerics*, 37(1):60–69.

Miller, F., Guilbaud, O., and Dette, H. (2007). Optimal designs for estimating the interesting part of a dose-effect curve. *Journal of Biopharmaceutical Statistics*, 17(6):1097–1115.

Müller, P. (1999). Simulation-based optimal design. *Bayesian Statistics*, 6:459–474.

Müller, P., Berry, D. A., Grieve, A. P., and Krams, M. (2006). A Bayesian decision-theoretic dose-finding trial. *Decision Analysis*, 3(4):197–207.

Müller, P. and Parmigiani, G. (1995). Optimal design via curve fitting of monte carlo experiments. *Journal of the American Statistical Association*, 90(432):1322–1330.

Müller, P., Sansó, B., and De Iorio, M. (2004). Optimal Bayesian design by inhomogeneous Markov chain simulation. *Journal of the American Statistical Association*, 99(467):788–798.

Myung, J. I. and Pitt, M. (2009). Optimal experimental design for model discrimination. *Psychological Review*, 116:499–518.

Napolitano, L., Park, P., Raghavendran, K., and Bartlett, R. (2009). Nonventilatory strategies for patients with life-threatening 2009 H1N1 influenza and severe respiratory failure. *Critical Care Medicine.*, 38(4 Suppl):74–90.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Ng, S. H. and Chick, S. E. (2004). Design of follow-up experiments for improving model discrimination and parameter estimation. *Naval Research Logistics*, 51:1129–1148.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

Ogungbenro, K. and Aarons, L. (2007). Design of population pharmacokinetic experiments using prior information. *Xenobiotica*, 37:1311–1330.

Ormerod, J. and Wand, M. (2010). Explaining variational approximations. *American Statistical Association*, 64(2):140–153.

Palmer, J. and Müller, P. (1998). Bayesian optimal design in population models for haematologic data. *Statistics in Medicine*, 17:1613–1622.

Paninski, L. (2005). Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17:1480–1507.

Pennanen, T. and Koivu, M. (2006). An adaptive importance sampling technique. In Niederreiter, H. and Talay, D., editors, *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 443–455. Springer Berlin, Heidelberg.

Pilz, J. (1991). *Bayesian estimation and experimental design in linear regregression models (2nd ed)*. Wiley, New York.

Pronzato, L. and Walter, E. (1985). Robust experiment design via stochastic approximation. *Mathematical Biosciences*, 75(1):103–120.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *The Annals of Statistics*, 19(3):1614–1625.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalised multilevel structural equation modelling. *Psychometrika*, 69(2):167–190.

Racine-Poon, A. (1985). A Bayesian approach to nonlinear random effects models. *Biometrics*, 41(4):1015–1023.

Retout, S. and Mentré, F. (2003). Further developments of the Fisher Information Matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics*, 13(2):209–227.

Rios Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer Verlag, New York.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407.

Ross, S. (1983). *Introduction to stochastic dynamic programming*. Academic Press.

Rossell, D. and Müller, P. (2013). Sequential stopping for high-throughput experiments. *Biostatistics*, 14(1):75–86.

Rossell, D., Müller, P., and Rosner, G. L. (2007). Screening designs for drug development. *Biostatistics*, 8(3):595–608.

Roth, P. (1965). *Design of Experiments for Discrimination Among Rival Models*. PhD thesis, Princeton University, New Jersey, USA.

Rubin, D., Hinton, S., and Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology*, 25(5):734–760.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Ryan, E., Drovandi, C. C., and Pettitt, A. N. (2014a). Fully Bayesian experimental design for pharmacokinetic studies. Technical report, Queensland University of Techonology.

Ryan, E., Drovandi, C. C., and Pettitt, A. N. (2014b). Simulation-based fully bayesian experimental design for mixed effects models. Technical report, Queensland University of Technology.

Ryan, E., Drovandi, C. C., Thompson, M., and Pettitt, A. N. (2014c). Towards Bayesian experimental design for nonlinear models that require a large number of sampling times. *Computational Statistics and Data Analysis*, 70:45–60.

Ryan, K. (2003). Estimating expected information gains for experimental designs with application to the random fatigue-limit model. *Journal of Computational and Graphical Statistics*, 12:585–603.

Saint-Marcoux, F., Knoop, C., Debord, J., Thiry, P., Rousseau, A., Estenne, M., and P., M. (2005). Pharmacokinetic study of tacrolimus in cystic fibrosis and non-cystic fibrosis lung transplant patients and design of bayesian estimators using limited sampling strategies. *Clinical Pharmacokinetics*, 44(12):1317–28.

Sansó, B. and Müller, P. (1997). Redesigning a network of rainfall stations. In *Case Studies in Bayesian Statistics IV*. Springer, New York.

Sepanski, J. H. and Kong, L. (2007). A family of generalized beta distributions for income. *arXiv:07104614v1 [stat.ME]*.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656.

Sheiner, L. and Beal, S. (1983). Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, 11(3):303–319.

Shekar, K., Roberts, J., Smith, M., Fung, Y., and Fraser, J. (2013). The ECMO PK project: an incremental research approach to advance understanding of the pharmacokinetic alterations and improve patient outcomes during extracorporeal membrane oxygenation. *BMC Anesthesiology*, 13:7.

Silvey, S. D. (1980). *Optimal design*. Chapman and Hall, London.

Sisson, S. A. and Fan, Y. (2011). *MCMC handbook*, chapter Likelihood-free Markov chain Monte Carlo, pages 313–335. Chapman & Hall.

Solonen, A., Haario, H., and Laine, M. (2012). Simulation-based optimal design using a response variance criterion. *Journal of Computational and Graphical Statistics*, 21(1):234–252.

Sorenson, D. and Gianola, D. (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer, New York.

Spall, J. C. (1998). An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins APL Technical Digest*, 19(4):482–492.

Spiegelhalter, D., Freedman, L., and Parmar, M. (1996). Bayesian approaches to randomize trials. In Berry, D. and Stangl, D., editors, *Bayesian Biostatistics*, pages 67–108. Dekker, New York.

Spiegelhalter, D. J. (2004). Incorporating Bayesian ideas into health-care evaluation. *Statistical Science*, 19(1):156–174.

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 157(3):357–416.

Stangl, D. and Berry, D. (1998). Bayesian statistics in medicine: Where are we and where should we be going? *The Indian Journal of Statistics*, 60:176–195.

Stigler, S. (1971). Optimal experimental designs for polynomial regression. *Journal of American Statistical Association*, 66:311–318.

Stroud, J., Müller, P., and Rosner, G. (2001). Optimal sampling times in population pharmacokinetic studies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):345–359.

Tan, F. and Berger, M. (1999). Optimal allocation of time points for the random effects models. *Communications in Statistics*, 28(2):517–540.

Tekle, F., Tan, F., and Berger, M. (2007). D-optimal cohort designs for linear mixed-effects models. *Statistics in Medicine*, 27(14):2586–2600.

Tian, Y. and Wang, D. (2009). Sequential Bayesian design for estimation of EDp. In *The 2nd International Conference on Biomedical Engineering and Informatics, 2009. BMEI'09.*

Toman, B. and Gastwirth, J. (1994). Efficienct robust experimental design and estimation using a data-based prior. *Statistical Sinica*, 4:603–615.

Tommasi, C. (2009). Optimal designs for both model discrimination and parameter estimation. *Journal of Statistical Planning and Inference*, 139:4123–4132.

Tsai, C. and Chaloner, K. (2002). *Case studies in Bayesian Statistics 5*, chapter Using Prior opinions to examine sample size in a clinical trial: two examples, pages 409–423. Springer-Verlag, New York.

Van Laarhoven, P. and Aarts, E. (1987). *Simulated Annealing: Theory and Applications*. Reider, Dordrecht.

Vanlier, J., Tiemann, C., Hilbers, P., and van Riel, N. (2014). Optimal experimental design for model selection in biochemical networks. *BMC Systems Biology*, 8:20.

Verdinelli, I. (1992). *Bayesian Statistics 4*, volume 4, chapter Advances in Bayesian Experimental Design (with Discussion), pages 467–481. Oxford University Press, Oxford.

Verdinelli, I. (2000). Bayesian design for the normal linear model with unknown error variance. *Biometrika*, 87:222–227.

Wakefield, J. (1994). An expected loss approach to the design of dosage regimens via sampling-based methods. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1):13–29.

Wand, M. and Jones, C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, 9(2):97 – 116.

Waterhouse, T. H., Eccleston, J. A., and Duffull, S. B. (2009). Optimal design criteria for discrimination and estimation in nonlinear models. *Journal of Biopharmaceutical Statistics*, 19:386–402.

Wathen, J. and Thall, P. (2008). Bayesian adaptive model selection for optimizing group sequential clinical trials. *Statistics in Medicine*, 27:5586–5604.

Whitehead, J. and Brunier, H. (1995). Bayesian decision procedures for dose determining experiments. *Statistics in Medicine*, 14:885–893.

Whitehead, J. and Williamson, D. (1998). Bayesian decision procedures based on logisic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics*, 8:445–467.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika*, 80:791–95.

Wolfson, L., Kadane, J., and Small, M. (1996). Expected utility as a policy making tool: an environmental health example. In Berry, D. and Stangl, D., editors, *Bayesian Biostatistics*, pages 261–277. Dekker, New York.

Woods, D. C., Lewis, S. M., Eccleston, J. A., and Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, 48:284–292.

Zhang, S. and Lee, M. (2010). Optimal experimental design for a class of bandit problems. *Journal of Mathematical Psychology*, 54(6):499–508.

Zidek, J., Sun, W., and Le, N. (2000). Designing and integrating composite networks for monitoring multivariate Gaussian pollution fields. *Applied Statistics*, 49:63–79.